Socioeconomic Inequalities in the Consequences of Gastrointestinal Infections

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Tanith Christine Rose

December 2017

Department of Public Health and Policy
University of Liverpool

Declaration

The work presented in this thesis is my own, except where work which has formed part of jointly-authored publications has been included (see Appendix 7 for publications). My contributions to these publications and those of the co-authors have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

Some of the work presented in Chapter 4 has formed two publications. The first has been published in the journal Systematic Reviews as 'Relationship between socioeconomic status and gastrointestinal infections in developed countries: a systematic review protocol' by Tanith Rose and Natalie Adams (students and joint first authors), David Taylor-Robinson, Benjamin Barr, Jeremy Hawker, Sarah O'Brien, Mara Violato and Margaret Whitehead. NA and TR wrote the protocol. DTR, BB, JH, SOB, MV and MW conceived the initial idea for the study, critically appraised the protocol and also contributed to its development by revising different versions. All authors approved the final version and take responsibility for its content.

The second publication has been submitted to the journal PLoS ONE as 'Relationship between socioeconomic status and gastrointestinal infections in developed countries: a systematic review and meta-analysis' by Natalie Adams and Tanith Rose (joint first authors), Jeremy Hawker, Mara Violato, Sarah O'Brien, Benjamin Barr, Victoria Howard, Margaret Whitehead, Ross Harris and David Taylor-Robinson. All authors contributed to the conception and design of the study. NA and TR performed the literature searches, and quality assessment of included studies. TR, NA and VH performed the data extraction. TR and NA performed the analyses with guidance from DTR, BB and RH, and all authors interpreted the data. NA and TR wrote the manuscript which was revised critically by all authors. All authors approved the final version of the manuscript.

Some of the findings presented in Chapter 5 have been published in the journal BMC Infectious Diseases as 'Socioeconomic status is associated with symptom severity and sickness absence in people with infectious intestinal disease in the UK' by Tanith Rose, Natalie Adams, Benjamin Barr, Jeremy Hawker, Sarah O'Brien, Mara Violato, Margaret Whitehead and David Taylor-Robinson. All authors contributed to the conception and design of the study. TR performed the analyses with guidance from DTR and BB. TR wrote the manuscript which was revised critically by DTR, BB, MV, JH, NA, SOB and MW. All authors approved the final version of the manuscript.

Dedication

I would like to dedicate this thesis to my parents, Michael and Christine, who have always believed in me, and have taught me to believe in myself.

Funding

This work was supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at the University of Liverpool in partnership with Public Health England, and in collaboration with the University of East Anglia, University of Oxford and the Institute of Food Research. Tanith Rose is based at the University of Liverpool. The views expressed are those of the author and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Acknowledgements

There are many people I would like to thank for their contributions to this research, and for supporting me during my studentship.

Firstly, I would like to thank my supervisors, David Taylor-Robinson, Ben Barr, Mara Violato, Sarah O'Brien, Jeremy Hawker and Margaret Whitehead. You are exceptional teachers and I have learnt a great deal from you over the past three years. Thank you for your guidance, patience and encouragement throughout this process. I feel immensely fortunate to have had such wonderful supervisors, who have made this experience a thoroughly enjoyable one.

I am also grateful to colleagues within the NIHR HPRU in Gastrointestinal Infections, the Department of Public Health and Policy at the University of Liverpool, and elsewhere, for their contributions and support, including Vivianne Buller, Kostas Daras, John Harris, Ross Harris, Victoria Howard, Chris Kypridemos, Sarah McGarrol, Daniel Pope, Sara Ronzi, Suzanne Rotheram and Peng Yin.

I would particularly like to thank Natalie for her support, encouragement and friendship throughout this journey.

Finally, a special thanks to my parents for their constant love and support, and for the many hours spent proofreading. Nick, thank you for helping me in so many ways, and for being incredibly accommodating, including relocating to the North West with me! Without you I could not have done this.

Thanks also to the NIHR HPRU in Gastrointestinal Infections for funding this research.

Abstract

Background: Gastrointestinal (GI) infections are very common and are associated with numerous adverse consequences for the individual, healthcare sector and economy as a whole. Relatively little is known about whether the consequences of having a GI infection are experienced by all members of society equally or whether certain groups are disproportionately affected. Some evidence suggests those of lower socioeconomic status (SES) are more likely to present to healthcare services with GI infections. This may reflect greater need amongst more disadvantaged groups, either due to increased risk of infection or disease severity. This thesis endeavours to expand current understanding, by comparing inequalities in the incidence of infection amongst cases occurring in the community and those presenting to healthcare services. In addition, it explores the extent of inequalities in disease severity, sickness absence and hospitalisation outcomes due to GI infections.

Methods: The framework of this thesis is based on theoretical knowledge of the mechanisms by which social stratification influences health outcomes. Three studies are presented. I begin by conducting a systematic literature review to examine inequalities in the risk of symptomatic GI infections in high income countries, using studies that have identified cases via healthcare records, laboratory notifications and population-based surveys. Narrative and meta-analytic methods are used to synthesise evidence and explore sources of statistical heterogeneity. I also analyse data collected in the UK-based Second Study of Infectious Intestinal Disease in the Community (IID2 study) to examine inequalities in self-reported symptom severity and sickness absence, amongst individuals with infectious intestinal disease (IID) aged $\geq$5 years. Regression modeling is used to investigate inequalities in these outcomes, whilst exploring the impact of several covariates such as age, sex, ethnicity, urban/rural residency and recent foreign travel. Finally, I perform an ecological analysis using routinely collected Hospital Episode Statistics data, to evaluate inequalities in emergency hospital admissions for IID and the duration of these admissions, across England. Stratified analyses for children and adults are performed, and the effects of several neighbourhood-level characteristics on inequalities in admissions are assessed.

Results: Firstly, the systematic literature review identified age as a statistically significant modifier of the association between SES and the risk of symptomatic GI infections. Children (aged <18 years) of lower SES, but not adults, had a greater risk of infection compared to their more affluent counterparts. Lower SES was also associated with higher risk of infection amongst studies that identified cases via hospitals, most of which analysed children only. Secondly, analysis of the IID2 study revealed that IID cases aged $\geq$5 years, of lower SES, were more likely to experience severe symptoms and be absent from work or school. The association between SES and sickness absence was largely explained statistically by greater symptom severity amongst the more disadvantaged cases. Thirdly, in English neighbourhoods, increasing deprivation was associated with increasing emergency hospital admission rates and admission duration for IID, for both adults and children. The social gradient in admission rates was partly explained statistically by geographical factors and the higher prevalence of long-term health problems in the more deprived neighbourhoods.

Conclusions: Important consequences of GI infections such as sickness absence, disease severity and emergency hospitalisation incur heavy burdens for individuals and societies. Evidence from this thesis suggests these adverse outcomes disproportionately affect socioeconomically disadvantaged groups. With this in mind, due consideration should be afforded to policies that address inequalities in the consequences of being ill with a GI infection, as well as current UK policies designed to reduce the risk of acquiring an infection.

Contents

Figures

Tables

Abbreviations

A&E	Accident and Emergency
ACSC	Ambulatory Care-Sensitive Condition
AIDS	Acquired Immunodeficiency Syndrome
ALSPAC	Avon Longitudinal Study of Pregnancy and Childhood
BMA	British Medical Association
CDRC	Consumer Data Research Centre
CI	Confidence Interval
CIP	Continuous Inpatient
E. coli	*Escherichia coli*
FSA	Food Standards Agency
GAM	Generalised Additive Model
GCSE	General Certificate of Secondary Education
GI	Gastrointestinal
GLM	Generalized Linear Model
GP	General Practice
HES	Hospital Episode Statistics
HIV	Human Immunodeficiency Virus
HR	Hazard Ratio
HSCIC	Health and Social Care Information Centre
HUS	Haemolytic Uraemic Syndrome
IBS	Irritable Bowel Syndrome
ICD-10	International Statistical Classification of Diseases and Related Health Problems 10th Revision
IID	Infectious Intestinal Disease
IID1	First Study of Infectious Intestinal Disease in the Community
IID2	Second Study of Infectious Intestinal Disease in the Community
ILRR	Integrated Longitudinal Research Resource
IMD	Index of Multiple Deprivation
IRR	Incident Rate Ratio
IVT	Intravenous Therapy
LQAT	Liverpool University Quality Assessment Tool
LSOA	Lower-Layer Super Output Area
MAR	Missing At Random
MCAR	Missing Completely At Random
MICE	Multivariate Imputation by Chained Equations
MID	Multiple Imputation then Deletion
MNAR	Missing Not At Random
MSOA	Middle-Layer Super Output Area
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR HPRU	National Institute for Health Research Health Protection Research Unit
NS-SEC	National Statistics Socioeconomic Classification
OA	Output Area

OECD	Organisation for Economic Co-operation and Development
OR	Odds Ratio
ORS	Oral Rehydration Salt
PCT	Primary Care Trust
PHE	Public Health England
QOF	Quality and Outcomes Framework
RR	Relative Risk
SD	Standard Deviation
SES	Socioeconomic Status
SOA	Super Output Area
STEC	Shiga toxin-producing *Escherichia coli*
SUS	Secondary Uses Service
UK	United Kingdom
UN	United Nations
UNICEF	United Nations Children's Fund
USA	United States of America
WHO	World Health Organization

Chapter 1

Introduction

1.1 RELEVANCE OF THE PROBLEM

The problem in context

This thesis explores socioeconomic inequalities in the consequences of symptomatic gastrointestinal (GI) infections (characterised by symptoms of diarrhoea and/or vomiting) in high income countries. Socioeconomic inequalities in health arise when systematic differences in health exist between groups which occupy unequal positions in society (Graham, 2007). For more than a century, researchers have compared measures of mortality and morbidity across socioeconomic groups (McKee and Pommerleu, 2005). They have found that health inequalities exist within and between countries across the world, reflecting differences in the opportunities to achieve good health between rich and poor (Morgan, 2006). In general, life expectancies and the prevalence of most diseases display a social gradient whereby the poorest in society experience greater levels of illness and premature death than those further up the socioeconomic scale (Wilkinson and Marmot, 2003). In the public health community, health inequalities that are socially produced are widely regarded as unfair and unjust (Whitehead and Dahlgren, 2007).

Levels of absolute poverty and adverse living conditions are important determinants of diarrhoea-related morbidity and mortality worldwide. Globally, child mortality rates due to infectious diarrhoea are disproportionately high in the world's poorest regions (United Nations Children's Fund [UNICEF], 2012). Approximately, 80% of child deaths due to diarrhoea occur in Africa and South Asia (UNICEF and World Health Organization [WHO], 2009). In high income countries, historical evidence also highlights a link between absolute poverty and morbidity due to GI infections. For example, in mid-19th century England and Wales, diarrhoea and other infectious diseases accounted for approximately 50% of all deaths (Omran, 2005). However, due to political change, improvements in living standards, hygiene and nutrition, and advancements in medicine and public health, Britain today has far lower levels of diarrhoea-related morbidity and mortality (Bambra, 2016; Omran, 2005; McKeown, Brown and Record, 1972).

Clearly, reducing the incidence and severe consequences of GI infections in low income countries should be an urgent global priority. In developed countries such as the United Kingdom (UK), however, where residents are able to benefit from relatively high standards of living, do GI infections still need to be prioritised within public health agendas?

Moreover, do the associations between poverty and adverse consequences of GI infections that have been observed historically in Britain, remain today?

The burden and consequences of GI infections

Over the course of at least two centuries, the UK has undergone an epidemiologic transition whereby infectious diseases have been displaced by chronic and degenerative diseases as leading causes of morbidity and mortality (Omran, 2005). Despite hopes held by leading members of the biomedical community in the 20th century that infectious diseases had been successfully conquered (Spellberg and Taylor-Blake, 2013; Fauci, 2001; Petersdorf, 1978), infectious diseases have persisted and are associated with a number of negative consequences.

In the UK, rates of symptomatic GI infections, otherwise known as infectious intestinal disease (IID), have increased in the community since the mid-1990s (Tam et al., 2012a). It is estimated that 17 million sporadic cases of IID occur every year (Tam et al., 2012a). That amounts to just over a quarter of the population experiencing an episode each year in the UK (Tam et al., 2012b). In addition to IID being extremely common, identified cases are frequently incapacitated due to their illness. Around half of those who experience IID report absence from work, school or daily activities (Tam et al., 2012b). This represents approximately eight million absences from school, and more than 11 million days lost from employment amongst people of working age per year (Food Standards Agency [FSA], 2016). These consequences may affect individuals differently depending on their socioeconomic circumstances.

The impact of IID on healthcare service utilisation is also sizeable. There are over one million general practice (GP) consultations for IID every year in the UK (Tam et al 2012a). Additionally, IID accounts for approximately 2% of all emergency hospital admissions in England (National Health Service [NHS] Digital, 2016). In 2015–16, this equated to 110,483 emergency hospital admissions with IID as the primary diagnosis. In a paediatric Accident and Emergency (A&E) department in Nottingham, England, 16% of all attendances over a one year period were due to diarrhoea (Armon et al., 2001). The burden on secondary care services is of particular importance, since most hospital admissions for IID are considered to be preventable events (Tian, Dixon and Gao, 2012).

The morbidity caused by IID is associated with considerable economic costs for both the individual and society. In 2010, the Food Standards Agency (FSA) estimated that foodborne illness (a subset of IID) alone cost the UK around £1.9 billion per annum (FSA, 2012). Whilst hospital admission for IID tends to occur relatively infrequently, the associated costs can be substantial. A report by The King's Fund estimated that emergency hospital admissions for dehydration and gastroenteritis cost the NHS in England just under £128 million in 2009–10 (Tian, Dixon and Gao, 2012). As mentioned, hospitalisation for IID is considered to be a preventable event, and thus hospital admissions for this condition represent expensive, yet potentially avoidable costs. Additionally, economic costs due to lost employment represent a large proportion of the overall economic burden due to IID (Roberts et al., 2003). A UK-based study found that the average cost to the health sector for an episode of acute gastroenteritis in children less than five years of age was £60, but the total cost to society was £176 when accounting for parental costs and the value of work time lost (Lorgelly et al., 2008).

Thus, whilst GI infections are usually self-limiting, the high frequency with which they occur in the community can amount to substantial overall societal costs. For the individual, GI infections can be associated with several adverse consequences such as sickness absence and hospitalisation, both of which may incur personal financial costs due to loss of work/income. The burdens and consequences of GI infections are clearly evident, however less is known about whether these burdens are experienced by all members of society equally or whether certain groups are disproportionately affected.

Current knowledge of inequalities in GI infections

Some studies conducted in high income countries, have found that those of lower socioeconomic status (SES) compared to high have higher rates of GP consultation (Phillips et al., 2011; Beale et al., 2010; Teschke et al., 2010; Quigley et al., 2006) and hospital admission due to GI infections (Biering-Sørensen et al., 2012; Lal et al., 2012; Wilking et al., 2012; Pockett et al., 2011; Moorin et al., 2010; Ma, El Khoury and Itzler, 2009; Özmert, Kilic and Yurdakök, 2008; Dennehy et al., 2006; Olowokure et al., 1999; Borgnolo et al., 1996). For example, in the West Midlands in the UK, hospital admission rates for young children with GI infections were found to be twice as high in the most deprived areas compared to the least (Olowokure et al., 1999). However, the mechanisms explaining these apparent health inequalities are poorly understood.

It could be that the apparent social gradient in healthcare use for GI infections seen in some studies reflects increased incidence of symptomatic infection amongst those of lower SES compared to high. If individuals of lower SES have a greater risk of symptomatic infection, they may also have a greater need for healthcare services and thus be more likely to present. A number of studies conducted in high income countries have measured inequalities in the risk of acquiring a symptomatic GI infection in the community, via population-based surveys. Yet the direction and magnitude of the association between SES and infection risk remain unclear. Several studies have observed an increased risk of GI infection amongst more socioeconomically disadvantaged groups (Beale et al., 2010; Özkan et al., 2007; Ludvigsson et al., 2006; Etiler, Velipasaoglu and Aktekin, 2004; Bozkurt, Özgür and Özçirpici, 2003; Bozkurt, Özgür and Özçirpici, 1999; Baker, Taylor and Henderson, 1998; Turkish Ministry of Health, 1995; Eaton-Evans and Dugdale, 1987). Whilst others have observed the opposite; an increased risk of infection amongst more socioeconomically advantaged groups (Adams et al., 2017; Pollard et al., 2014; Van Cauteren et al., 2012; Scallan et al., 2004; Herikstad et al., 2002; De Wit et al., 2001a; Fein, Lin and Levy, 1995). Studies that have measured this association have been conducted in a range of high income countries, using different measures of SES and study designs, whilst adjusting for a variety of potential confounding factors. Thus, whilst there are many sources of clinical and methodological heterogeneity amongst studies that have measured inequalities in the incidence of GI infections, it is unknown to what extent these sources contribute to the observed contrasting results.

Alternatively, inequalities in healthcare utilisation for GI infections might be explained by increased levels of disease severity amongst lower socioeconomic groups. This hypothesis is supported by evidence from studies that have found disease severity to be a strong predictor of GP presentation for IID (Doorduyn, Van Pelt and Havelaar, 2012; Van Cauteren et al., 2012; Adlam et al., 2011; Scallan et al., 2006; Tam, Rodrigues and O'Brien, 2003; De Wit et al., 2001b), and hospital admission is in itself a severe consequence of having a GI infection. Some studies have found inequalities in measures of IID severity such as the duration of illness, however these are sparse in number and have exclusively focused on paediatric populations (Ma, El Khoury and Itzler, 2009; Baker, Taylor and Henderson, 1998; Conway, Phillips and Panday, 1990). In the UK, the extent of inequalities in IID symptom severity amongst individuals of all ages has not previously been investigated.

Finally, only three ecological studies have investigated inequalities in IID-related hospital admissions in the UK. Of these, two found evidence of a socio-spatial gradient in admission rates (Pockett et al., 2011; Olowokure et al., 1999), and one found no statistically significant

relationship between deprivation and admission rates (Kyle et al., 2011). Two of the studies aggregated data over large geographical areas, and analysed paediatric populations only (Pockett et al., 2011; Kyle et al., 2011), and Olowokure et al. (1999) analysed data collected over 20 years ago.

Hence, there are several gaps in the literature in relation to inequalities in GI infections. The work in this thesis endeavours to address some of these gaps, to enhance current understanding of inequalities in the consequences of GI infections.

1.2 RATIONALE FOR THIS RESEARCH

As outlined above, GI infections are very common and are associated with numerous adverse consequences for the individual, healthcare sector and economy as a whole. There is some evidence that inequalities in healthcare presentation for GI infections are apparent. Little is known, however, about the mechanisms that might explain these inequalities. Contributing factors may include differential risk of infection, or differential disease severity across socioeconomic groups. Assessing and gaining a better understanding of the relative importance of these possible explanations is essential to inform the development of effective interventions and policies to tackle any inequalities observed. Research is required to address the gaps in the knowledge base that have been highlighted and to offer new insights. This thesis seeks to address some of these gaps.

The theoretical basis of this thesis is guided by Diderichsen's model of the mechanisms generating socioeconomic inequalities in health (Diderichsen, Evans and Whitehead, 2001). Diderichsen's model outlines the potential pathways and mechanisms involved in the generation of health inequalities at the individual and population level, whereby an individual's social position determines their exposure and vulnerability to disease, and the consequences they experience as a result. The studies that I present in this thesis investigate the mechanisms of health inequality set forth in Diderichsen's model, in the context of GI infections (Figure 1.1). I endeavour to expand current understanding, by comparing inequalities in the incidence of infection amongst cases occurring in the community and presenting to healthcare services (Study 1), and by measuring the extent of inequalities in disease severity, sickness absence and hospitalisation as consequences of GI infections (Studies 2 and 3).

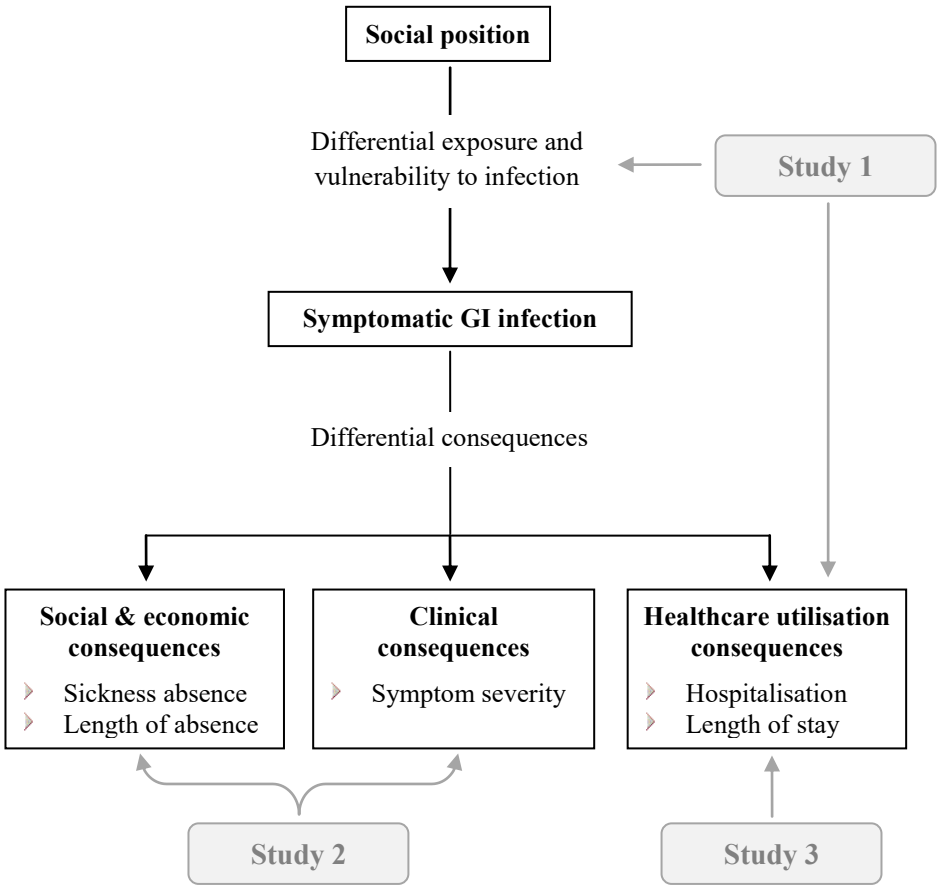
1.3 AIMS AND OBJECTIVES

The overarching aim of this thesis is to assess the extent of socioeconomic inequalities in various consequences of GI infections, and to explore possible explanations for any inequalities identified.

The objectives are:

1) To systematically review current evidence on the relationship between SES and the incidence of symptomatic GI infections in high income countries, using studies that have identified cases via healthcare records, laboratory notifications and population-based surveys (Study 1).
2) To investigate the association between SES and self-reported IID symptom severity and sickness absence, using data collected in the Second Study of Infectious Intestinal Disease in the Community (IID2 study) in the UK (Study 2).
3) To assess the impact of neighbourhood income deprivation on emergency hospital admission rates for IID and the duration of these admissions in England, using Hospital Episode Statistics (HES) data (Study 3).
4) To reflect on the empirical findings of the three studies above, and draw out implications for policy.

Figure 1.1 Overview of the studies in this thesis using Diderichsen's model of the mechanisms of health inequality as a framework



Source: Adapted from Diderichsen, Evans and Whitehead (2001)

1.4 STRUCTURE OF THIS THESIS

The structure of this thesis and the following chapters keep the aim and objectives central to the inquiry. As demonstrated in Figure 1.1, I endeavour to gain a better understanding of inequalities in the incidence of GI infection in Study 1. The results of this study may help to explain any inequalities identified in Study 3, which investigates the social patterning of secondary healthcare use as a consequence of having a GI infection. Additionally, Study 2 explores inequalities in outcomes such as symptom severity and sickness absence due to IID.

The chapters of the thesis are outlined as follows:

- Chapter 2, the Literature Review, explores the existing literature and background information relevant to this thesis. An overview of GI infections and health inequalities

is provided, and the literature investigating inequalities in GI infections is described and critiqued. Gaps in current knowledge are identified and discussed.

- Chapter 3, the Methods, specifies in detail the methods utilised in each of the three studies conducted in this thesis. The rationale behind the choice of each method is explained.

- Chapter 4, the results of Study 1, outlines and discusses the results of a systematic literature review and meta-analysis, which investigates the association between SES and symptomatic GI infection risk in high income countries (addressing objective 1).

- Chapter 5, the results of Study 2, outlines and discusses the results of a cross-sectional analysis of data collected in the population-based UK IID2 study, which explores the association between SES and self-reported IID symptom severity and sickness absence due to IID (addressing objective 2).

- Chapter 6, the results of Study 3, outlines and discusses the results of a cross-sectional ecological analysis of routinely collected HES data in England, which examines the association between deprivation and emergency hospital admission rates and admission duration for IID (addressing objective 3).

- Chapter 7, the Discussion, provides an overview of the results of the three studies presented in this thesis, and the ways in which the work has made a unique and original contribution to the literature. In this chapter, I evaluate and synthesise the results of the three studies in the context of other relevant literature, to develop greater insight into the social patterning of the consequences of GI infections. Unanswered questions and future research recommendations are considered, as well as the limitations of the research and implications for policy and practice (addressing objective 4).

Finally, the appendices for Chapters 4, 5 and 6 (Appendices 4, 5 and 6, respectively) contain additional information and analyses to support the main findings, and Appendix 7 details the publications from this thesis.

Chapter 2

Literature review

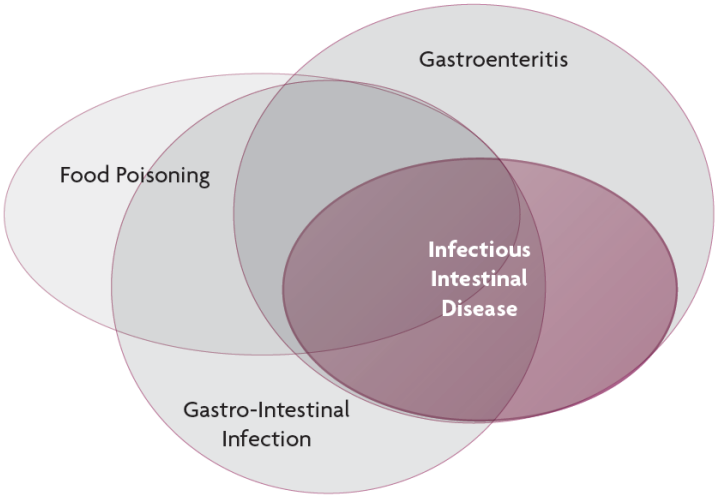
This chapter explores the literature and background information that is relevant to this thesis. Literature pertaining to GI infections and health inequalities is outlined first, before focussing on an overview of the literature investigating socioeconomic inequalities in GI infections. Gaps in current knowledge are identified.

2.1 GI INFECTIONS

Definitions

To begin, it is useful to differentiate between some of the terms related to GI infections that have been used in the literature. GI infection, IID, food poisoning and gastroenteritis are terms that are used somewhat inter-changeably, but have slightly different definitions. Gastroenteritis describes inflammation of the mucus membrane lining the stomach and intestines and is characterised by symptoms such as diarrhoea and vomiting (Collin, 1998). Certain conditions, such as Crohn's disease and irritable bowel syndrome, may cause gastroenteritis, but are non-infectious in nature (Tam et al., 2012b). Likewise, food poisoning may also be caused by non-infectious agents, e.g. mycotoxins or mercury, but instances of food poisoning may or may not give rise to the symptoms of gastroenteritis (Tam et al., 2012b). GI infections and IID are similar in that they are both caused by infectious agents, however IID is always characterised by symptoms of gastroenteritis (Tam et al., 2012b). Figure 2.1 shows the inter-related nature of the terms.

Figure 2.1 Relationship between GI infections and related terms



Source: Tam et al. (2012b)

Causes, transmission and prevention

Infections of the GI tract can be caused by a variety of pathogenic agents. If not destroyed by the natural defences of the GI system, enteric pathogens such as bacteria, viruses and parasites, can cause disease by adhering to the gut mucosa, invading enterocytes or producing enterotoxins (Hamer and Gorbach, 2010; Bannister, Gillespie and Jones, 2009). These microorganisms, can thus compromise the absorptive properties of the bowel leading to diarrhoea and potential dehydration (Bannister, Gillespie and Jones, 2009).

Humans can become infected via various modes of transmission, which vary by pathogen. Many infectious agents are commonly transmitted via the faecal-oral route from human or animal reservoirs (Hawker et al., 2008). Transmission pathways can usually be described as direct (via contact or droplets); or indirect (via food/water vehicles, vectors or airborne droplets) (O'Brien and Halder, 2007). Many GI infections are transmitted to humans via contaminated food and water (Lamps, 2009). Viral pathogens are commonly transmitted directly via person-to-person transmission (Wikswo and Hall, 2012; Dennehy, 2000). A description of common GI pathogens is presented in Table 2.1.

The probability of infection depends on a variety of factors related to the host, the infecting pathogen and the environment (O'Brien and Halder, 2007). For example, the number of cells required to infect a host (known as the infectious dose) varies by pathogen, with some such as *Giardia* requiring approximately 10 cells to start an infection, and others such as *Vibrio cholerae* requiring around 10^3 to 10^8 cells (Leggett, Cornwallis and West, 2012; Schmid-Hempel and Frank, 2007). Additionally, factors related to the environment, such as the dose to which the host is exposed, and host-related factors such as immunocompetence may influence the likelihood of infection (O'Brien and Halder, 2007). Hosts previously exposed to certain pathogens may also develop a certain level of immunity to future disease (Janssen et al., 2008).

Several relatively simple precautions can be implemented whilst storing and preparing food to prevent many foodborne GI infections. These include storing perishable food at safe temperatures (below 5°C), cooking food thoroughly, and preparing raw meat separately to foods that will not be cooked (WHO, 2006). Hand washing with soap after using the toilet and before preparing food and eating can also prevent many GI infections (Centers for Disease Control and Prevention, 2015). Additionally, a rotavirus vaccine is provided routinely in the UK for infants aged two and three months, and since its introduction has prevented >70% of reported cases (Table 2.1) (Karafillakis, Hassounah and Atchison, 2015).

Table 2.1 Common pathogens that can infect the GI tract

Campylobacter

Campylobacter bacteria are the most commonly reported cause of food poisoning in the UK (Public Health England [PHE], 2017a). In 2015, there were 96.2 laboratory reports per 100,000 population in England and Wales (PHE, 2017a). Campylobacteriosis in humans is predominantly caused by the species *Campylobacter jejuni*, and less often *Campylobacter coli* (Janssen et al., 2008). Whilst *Campylobacter* species are widely distributed in most warm-blooded animals, the majority of cases in the UK come from contaminated poultry (WHO, 2016a; FSA, no date). The main route of transmission is considered to be foodborne, and the infectious dose can be low, with around 500 cells required to cause infection in some individuals (Leggett, Cornwallis and West, 2012; Black et al., 1988). Symptoms often include severe abdominal pain, and watery or bloody diarrhoea (Janssen et al., 2008). A small number of cases experience post-infectious complications, such as reactive arthritis (<10% of cases) and Guillain-Barré syndrome (<2% of cases) (Esan et al., 2017; Keithlin et al., 2014).

Shiga toxin-producing *Escherichia coli* (STEC)

STEC is a pathogenic *Escherichia coli* (*E. coli*) strain that produces Shiga toxins (WHO, 2016b). The most frequently reported STEC strain to cause illness in the UK is *E. coli* O157 (PHE, 2014). Infection is relatively rare; in 2012 there were 795 cases of *E. coli* O157 reported to laboratories in England and Wales (PHE, 2014). In the UK, the main reservoir for STEC is cattle and other ruminants, and transmission can occur via contaminated food or water, or via environmental exposure to animals or their faeces (Adams et al., 2016). STEC has a very low infectious dose (around 10 cells can cause infection), and disease severity can range from mild to severe (PHE, 2014; Lim, Yoon and Hovde, 2010). Life-threatening post-infectious complications such as haemolytic uraemic syndrome (HUS) can develop in around 10–15% of cases (WHO, 2016b; Tarr, Gordon and Chandler, 2005).

***Salmonella* (non-typhoidal)**

There are thousands of different serotypes of *Salmonella* bacteria. The two most commonly identified in the UK are *Salmonella enterica* serotype Enteritidis, and *Salmonella enterica* serotype Typhimurium (WHO, 2016c). In 2015, there were 14.8 laboratory reports of *Salmonella* per 100,000 population (PHE, 2017b). *Salmonella* bacteria are found in the GI tracts of a wide variety of domestic and wild animals, and transmission is predominantly via contaminated food (WHO, 2016c). The infectious dose is fairly high, although low doses may still produce illness (PHLS Advisory Committee on Gastrointestinal Infections, 2004). Amongst foodborne bacterial pathogens, *Salmonella* has been identified as a leading cause of mortality in the United States of America (USA) and hospitalisations in the UK (O'Brien et al., 2016; Scallan et al., 2011).

Shigella

Shigella, also known as bacillary dysentery, has four species: *Shigella dysenteriae*; *Shigella flexneri*; *Shigella sonnei* and *Shigella boydii* (PHE, 2017c). There were 3.6 laboratory recorded *Shigella* infections per 100,000 population in 2015 in England and Wales (PHE, 2017c). The majority of reported cases of shigellosis in the UK are travel-related (PHE, 2017c). However, in recent years an increase in UK-acquired *Shigella flexneri* cases have been observed, which is likely due to transmission amongst men who have sex with men (Borg et al., 2012). The human GI tract is a reservoir of *Shigella* infection, and person-to-person transmission is common (PHLS Advisory Committee on Gastrointestinal Infections, 2004). Shigellosis is often associated with bloody diarrhoea and *Shigella dysenteriae* serotype 1 has been associated with the development of HUS, a life-threatening post-infectious complication (PHLS Advisory Committee on Gastrointestinal Infections, 2004).

Rotavirus

In the UK, rotavirus is a common cause of IID in infants and young children (PHE, 2017d). Following the introduction of a rotavirus vaccine to the childhood immunisation programme in 2013, laboratory reports of rotavirus infections have dramatically declined in England and Wales, from 26.3 to 7.7 reports per 100,000 population in 2013 and 2014, respectively (PHE, 2017d). The human GI tract is a reservoir of rotavirus infection, person-to-person transmission is very common and the infectious dose is low (Glass et al., 2006; PHLS Advisory Committee on Gastrointestinal Infections, 2004). There is evidence to suggest that among young children under six years of age, rotavirus infectious are likely to produce more severe symptoms compared to other viral GI infections (Iturriza Gómara et al., 2008).

Norovirus

Norovirus has been identified as the commonest cause of IID in the UK using population-based studies (Tam et al., 2012c). However, since few cases visit their GP for their illness, and thus have a stool sample taken, laboratory records do not reflect the high incidence of norovirus infections in the community (PHE, 2017e; Tam et al., 2012c). In 2015, there were 11.6 laboratory reported cases per 100,000 population in England and Wales (PHE, 2017e). The human GI tract is a reservoir of norovirus infection, and person-to-person transmission is very common (PHLS Advisory Committee on Gastrointestinal Infections, 2004). The infectious dose is low, and symptoms are usually mild (PHE, 2017e).

Cryptosporidium

Cryptosporidiosis is predominantly caused by the parasites *Cryptosporidium hominis* and *Cryptosporidium parvum* (Eckert, 2005). In 2015, there were 9 per 100,000 laboratory reports of *Cryptosporidium* infections in England and Wales (PHE, 2017f). *Cryptosporidium hominis* is transmitted within the human population, and humans can also acquire zoonotic infections from contact with infected animals, particularly cattle and sheep (Eckert, 2005). Transmission can occur directly and also via contaminated food or drinking water, and from swimming in contaminated water (Smith et al., 2010). The parasites are resistant to chlorine and thus water suppliers remove the parasites by filtration methods and routinely monitor the treated water (Smith et al., 2010). Symptoms can range from mild to severe, which can depend on the immune status of the host (PHLS Advisory Committee on Gastrointestinal Infections, 2004).

Giardia

Giardiasis is caused by the parasite *Giardia intestinalis* which is also known as *Giardia duodenalis* and *Giardia lamblia* (Eckert, 2005). In the UK, many cases are associated with foreign travel (PHE, 2017g). In 2015, there were 7.6 per 100,000 laboratory reports of *Giardia* infections in England and Wales (PHE, 2017g). The GI tracts of humans are considered an important reservoir, and it is unclear whether animals may also serve as sources of infection for humans, but if so this is probably rare (PHLS Advisory Committee on Gastrointestinal Infections, 2004). Transmission can occur person-to-person, via contaminated food or drinking water, and from swimming in contaminated water (Eckert, 2005). *Giardia intestinalis* is resistant to chlorine, and the infectious dose is very low (Leggett, Cornwallis and West, 2012; PHLS Advisory Committee on Gastrointestinal Infections, 2004). Clinical presentation can range from asymptomatic to severe disease which can depend on the host's ability to eliminate the parasite (Eckert, 2005).

Diagnosing GI infections

Clinicians and sometimes microbiologists are responsible for diagnosing GI infections (Lamps, 2009). In practice, a diagnosis of gastroenteritis is usually made on the basis of clinical symptoms and signs, and many cases are managed without determining a microbial diagnosis (National Institute for Health and Care Excellence [NICE], 2015). Acute gastroenteritis may be suspected if patients present with three or more loose stools in a 24-hour period (NICE, 2015). Differential diagnoses, such as chronic idiopathic inflammatory bowel disease or diarrhoea caused by certain drugs, may also be considered (NICE, 2015; Lamps, 2009). Current recommendations in the UK advise that stool microbiological investigations to determine the infecting pathogen are not required for the majority of cases (NICE, 2015). Stool cultures may be required if patients are immunocompromised, have a history of recent hospitalisation or exotic foreign travel, or if diarrhoea is persistent or severe and there is uncertainty about the diagnosis (NICE, 2015).

Clinical management

Professional medical advice is not sought by the majority of people with a GI infection in the UK, and recovery usually occurs within days (National Collaborating Centre for Women's and Children's Health, 2009; Wheeler et al., 1999). The NHS provides online advice for individuals with acute diarrhoea, and recommends increasing fluid intake, resting, eating plain foods and taking oral rehydration salt (ORS) solution if dehydration occurs (NHS Choices, 2016). ORS typically contains a mixture of glucose and sodium chloride, which aid intestinal salt and water absorption (National Collaborating Centre for Women's and Children's Health, 2009).

Individuals are advised to seek medical advice if their symptoms are severe, have not improved, or if they have a serious underlying medical condition (NHS Choices, 2016). Treatment for gastroenteritis is mostly supportive and is directed at preventing and treating dehydration (Burkhart, 1999). It is recommended that patients presenting to primary care should be given rehydration advice (NICE, 2015). Children without clinical dehydration but who are at increased risk of dehydration should be offered ORS solution, as should children with clinical dehydration who can be managed at home (NICE, 2015). For healthy adults who are able to maintain their fluid intake, ORS solution is not indicated, however ORS solution may be considered for adults at increased risk of poor outcomes, such as the elderly and those with comorbidities (NICE, 2015; Farthing et al., 2012; Wingate et al., 2001).

Emergency hospital admission is indicated if individuals have signs of shock or severe dehydration (NICE, 2015). Clinicians may also consider hospital admission when symptoms are severe and persistent, if children are younger than six months and adults are aged 60 years or over, or if patients have coexisting medical conditions (NICE, 2015). Treatment in secondary care settings often involves the use of intravenous therapy (IVT), which can rapidly and effectively reverse hypovolemic shock due to dehydration (Bellemare et al., 2004). Whilst IVT is indicated for children with shock, a Cochrane systematic review found that ORS solution was just as effective as IVT for the treatment of children with dehydration (National Collaborating Centre for Women's and Children's Health, 2009; Hartling et al., 2006). There is evidence to suggest however that IVT is overused in some high income countries (Chow, Leung and Hon, 2010).

During periods of acute GI infection, individuals are advised to take certain precautions to minimise the spread of the infection. For example, frequent hand-washing, cleaning toilet surfaces daily, and washing soiled clothing, bedding and towels at high temperatures, are all advised (PHLS Advisory Committee on Gastrointestinal Infections, 2004). In most situations, absence from work or school is recommended until at least 48-hours after symptoms recede (NICE, 2015). Children and other groups that pose an increased risk of spreading infection, that have gastroenteritis caused by STEC or other pathogens that have potentially serious sequelae, may be excluded for longer (PHLS Advisory Committee on Gastrointestinal Infections, 2004).

2.2 INCIDENCE AND MORTALITY

Worldwide, GI infections confer a significant proportion of overall morbidity and mortality in the population. In 2010, estimates suggested that there were around 4.6 billion cases (95% confidence interval [CI] 3.5–6.5) of foodborne diarrhoea, resulting in 1.6 million deaths (95% CI 1.3–1.9) globally (Pires et al., 2015). Around 40% of these cases and 43% of deaths occurred in children less than five years of age (Pires et al., 2015). Presently, diarrhoeal disease is the second leading cause of death worldwide in children under five (WHO, 2013).

Mortality rates due to diarrhoea vary markedly throughout the regions of the world. Internationally, child mortality rates due to diarrhoea are disproportionately high in the world's poorest regions (UNICEF, 2012). Approximately, 80% of child deaths due to

diarrhoea occur in Africa and South Asia (UNICEF and WHO, 2009). This relates to poor sanitation, lack of clean drinking water and unavailability of treatments (Farthing et al., 2012). In the UK however, diarrhoea-related mortality is rare. In 2015, there were 1358 recorded deaths with IID as the underlying cause in England and Wales (Table 2.2). The majority of these deaths occurred in the elderly population.

Table 2.2 Deaths due to IID in England and Wales

	Year		
Age group (years)	**2013**	**2014**	**2015**
0–4	8	6	7
5–14	2	2	1
15–44	18	15	13
45–64	80	64	62
65–74	141	143	162
75+	1327	1038	1113
Total (all ages)	**1576**	**1268**	**1358**

Deaths with underlying cause ICD-10 codes A00–A09 Intestinal infectious diseases
Data source: Nomis, Office for National Statistics

Despite the low levels of mortality due to IID in the UK, infections occur frequently in the population. Two large-scale population-based studies have been conducted in the UK to determine the incidence of IID in the community. The First Study of Infectious Intestinal Disease in the Community (IID1) was conducted in England in 1993–6, and the second study (IID2) was conducted across the UK in 2008–9. Both studies used similar methods and IID case definitions. The studies contained a prospective cohort study to estimate the incidence of IID in the community, and a GP presentation study to assess the incidence of IID presenting to primary care (Tam et al., 2012a). Results from the IID1 study indicated that in 1993–6 one in five individuals experienced an episode of IID per year, a crude incidence rate of 194 cases per 1000 person-years (95% CI 181–208) (Wheeler et al., 1999). Of those who experienced an episode, one in six consulted a GP for their illness, a consultation rate of 33.1 per 1000 person-years (95% CI 29.4–37.5) (Wheeler et al., 1999).

In contrast, the IID2 study in 2008–9 reported a crude IID incidence rate of 258 cases per 1000 person-years (95% CI 244–273) (the age and sex standardised rate was 274 cases per 1000 person-years; 95% CI 254–296), and a GP consultation rate of 17.7 per 1000 person-years (95% CI 14.4–21.8) (Tam et al., 2012a; Tam et al., 2012b). Thus, over the fifteen year

period between these similar studies, it appears that rates of IID increased whilst GP consultation rates decreased. Increased levels of self-management of IID or the introduction of telephone advice services such as NHS Direct have been suggested as possible explanations for the decrease in IID consultation rates. However, use of telephone advice services was low in the IID2 study, with less than 2% of individuals with IID reporting having contacted NHS Direct or NHS24 for their illness (Tam et al., 2012b). Speculatively, increased use of the internet for health information may have also contributed to the reduction in GP consultations for IID.

Interestingly, it appears that incidence estimates vary noticeably depending on the study design used. The IID2 study also included a telephone-based survey, designed to estimate the incidence of IID in the community retrospectively, using the same case definition as other IID2 studies. Participants were asked to recall whether they had experienced an episode of IID, either in the previous seven days, or in the previous 28 days (Viviani et al., 2016). The age and sex standardised IID incidence rate in the 7-day recall group was 1530 cases per 1000 person-years (95% CI 1135–2113), compared to 533 cases per 1000 person-years (95% CI 377–778) for the 28-day recall group (Viviani et al., 2016). Thus IID incidence in the 7-day recall group was almost treble that of the 28-day recall group. The authors suggested that participants may have forgotten milder illness in the 28-day recall group. Alternatively, IID cases in the 7-day recall group may have been more likely to be at home to receive the telephone call due to their illness (Viviani et al., 2016). The incidence estimates from both recall groups are also much higher than those reported in the IID2 prospective cohort study. Possible explanations for this might include reporting fatigue amongst prospectively followed cohort participants (Tam et al., 2012a). This seems plausible, since incident cases identified in the cohort study were asked to complete a symptom questionnaire and submit a stool sample, which may have deterred participants from reporting their illness.

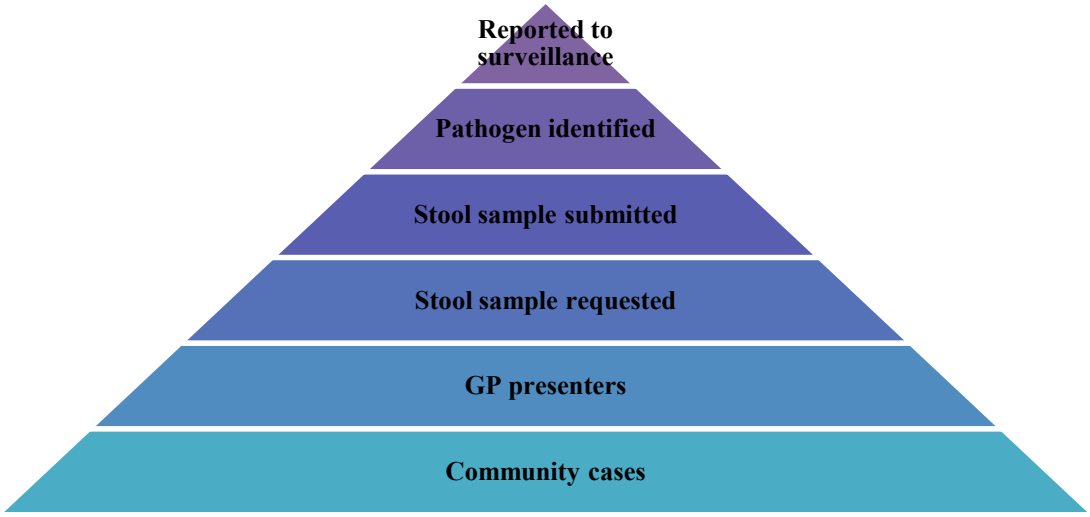
Stool samples were requested from all cases identified in the IID1 and IID2 cohort and GP presentation studies. In the IID2 prospective cohort study, a pathogen was identified in only 40% of the stool samples submitted (Tam et al., 2012c). Viruses were the most frequently detected organisms and norovirus was the leading pathogen, identified in 16.5% of the samples (Tam et al., 2012c). The most common bacterial pathogen was *Campylobacter* identified in 4.6% of the samples, and other bacterial and protozoal pathogens were detected infrequently (<2%) (Tam et al., 2012c). Somewhat similar proportions were observed amongst cases who consulted a GP for their illness, although consultations for norovirus infections were relatively low. *Campylobacter* (13%) and norovirus (12.4%) were the most frequently detected pathogens amongst GP presenting cases (Tam et al., 2012c).

Surveillance in the UK and the reporting pyramid

In the UK today, public health bodies such as PHE and Health Protection Scotland are responsible for monitoring the incidence of certain GI infections as directed by law. The national surveillance of GI infections can assist with the identification of trends, and can highlight potential emerging issues and outbreaks (Health Protection Agency, 2011). All laboratories in England performing a primary diagnostic role have a statutory duty to report notifiable infectious organisms to PHE (GOV.UK, 2010). Notifiable infectious organisms include pathogens that have the potential to cause food poisoning, such as *Salmonella*, *Campylobacter*, *Shigella*, *Cryptosporidium*, *Giardia*, *Listeria* and *E. coli* O157 (GOV.UK, 2010). These notifications are collated and published by PHE on a weekly basis (GOV.UK, 2014). Outbreaks of foodborne and non-foodborne GI infections (suspected when two or more cases with the same infection are linked to the same source, or when the observed number of cases exceeds the expected number and the same source is suspected) are also monitored by PHE (PHE, 2013).

An issue when monitoring trends in the incidence of GI infections using laboratory surveillance systems, is that laboratory notifications capture a small proportion of the disease burden in the community. This is because several steps are required for a case to be notified to national surveillance, i.e. individuals identified as 'cases' will have presented to primary care, had a stool sample requested and had a pathogen identified in their stool. Under-ascertainment of cases occurs at each step (Tam et al., 2012b). This is sometimes referred to as the reporting or surveillance pyramid (Figure 2.2).

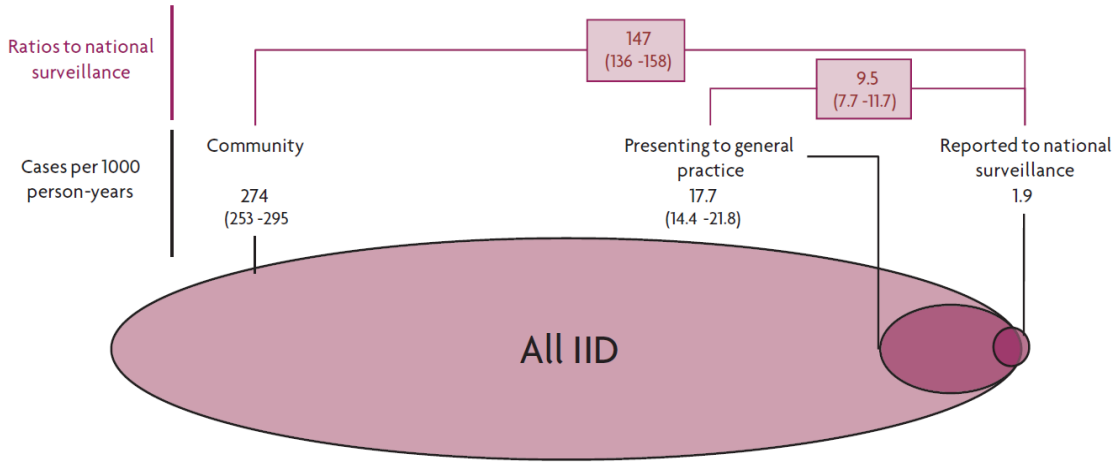
Figure 2.2 Surveillance/reporting pyramid



Source: Adapted from Tam et al. (2012b)

An objective of the IID1 and IID2 studies was to provide information to calibrate national laboratory surveillance systems by investigating the ratio of cases reported to national surveillance, to cases occurring in the community (Tam et al., 2012a). The IID2 study found that for every case reported to national surveillance, approximately 150 cases occur in the community (Tam et al., 2012b). The IID2 study team reported their results as ellipses to illustrate that a certain proportion of cases reported to national surveillance originate from hospitals and other institutions (Figure 2.3) (Tam et al., 2012b).

Figure 2.3 Reporting ratios from IID2 study



Source: Tam et al. (2012b)

In addition to laboratory notifications, symptom-based surveillance, known as syndromic surveillance, can also be used to monitor trends in the incidence of GI infections in the community. Reports of patients presenting to various healthcare providers (such as NHS Direct/111, GPs and emergency departments) with diarrhoea or vomiting symptoms are collected and analysed by PHE (Todkill et al., 2016). This approach has benefits in that monitoring is conducted in real-time to enable the early detection of trends, however a potential limitation of this approach is that it lacks specificity, particularly in relation to detecting outbreaks (Todkill et al., 2016). Additionally, when outbreaks are reported in the media, increases in NHS Direct/111 calls or GP presentations for diarrhoea might reflect changes in healthcare-seeking behaviour rather than increased disease burden in the community (Elliot et al., 2016).

2.3 CONSEQUENCES OF GI INFECTIONS

The consequences of GI infections are varied and can be wide reaching at the individual-level and wider societal-level. The following section is divided into: disease severity and clinical consequences; healthcare utilisation consequences; and social and economic consequences of GI infections.

Disease severity and clinical consequences

Several clinical complications can result following an episode of IID (a selection of which is detailed in Table 2.3), although these tend to occur infrequently. An exception is dehydration, a common complication (BMJ Best Practice, 2016) which can require hospitalisation if severe. Dehydration occurs when water and electrolytes (such as sodium, chloride, potassium and bicarbonate) lost via diarrhoea and/or vomiting, are not adequately replenished (WHO, 2005). In children, mild dehydration occurs when 5% of a child's body weight is lost (Elliott, 2007). Severe dehydration occurs when weight loss is 10% or more, and if not corrected, hypovolemic shock and death can result although this is rare in high income countries (BMJ Best Practice, 2016; Elliott, 2007). The severity of dehydration can depend on the duration of symptoms (Farthing et al., 1996).

The duration of diarrhoea may thus provide a good indication of disease severity and the extent of dehydration. An episode of acute gastroenteritis typically lasts around 5–7 days (Giannattasio, Guarino and Lo Vecchio, 2016). Prolonged diarrhoea, otherwise known as acute-protracted diarrhoea, is defined as diarrhoea with acute onset lasting from 7–14 days, and can be caused by persistent infections or complications of infections such as post-infectious intestinal damage (Giannattasio, Guarino and Lo Vecchio, 2016). The definition of diarrhoea lasting 14 days or more is less clear-cut, and can be termed persistent diarrhoea, chronic diarrhoea or intractable diarrhoea, and these illnesses may not necessarily be due to a GI infection (Giannattasio, Guarino and Lo Vecchio, 2016). Intractable diarrhoea has varied definitions and aetiologies, but represents diarrhoea that persists despite extensive hospital therapy, and is unusual in high income countries (Giannattasio, Guarino and Lo Vecchio, 2016; Hizarcioglu-Gulsen et al., 2014; Guarino et al., 1995).

Table 2.3 Clinical complications of GI infections

Reactive arthritis
Reactive arthritis is an inflammatory arthritis which can develop following a GI infection, usually within 1–2 weeks of the infection, although a definitive time period from infection to onset has not been established (Ajene, Fischer Walker and Black, 2013; Hannu, 2011). The bacterial pathogens *Salmonella*, *Shigella*, *Campylobacter*, *Yersinia* and *E. coli* have been associated with the development of reactive arthritis (Hannu, 2011). There exists no universally agreed upon diagnostic criteria and thus incidence estimates vary widely (Hannu, 2011). In a systematic review, reactive arthritis incidence ranged from 0–16%, 0.1–29%, and 0–12% among *Campylobacter*, *Salmonella* and *Shigella* infections respectively (Ajene, Fischer Walker and Black, 2013). In a more recent systematic review, the majority of studies that were identified, found that <10% of individuals developed reactive arthritis following *Campylobacter* or non-typhoidal *Salmonella* infections (Esan et al., 2017).
Post-infectious irritable bowel syndrome (IBS)
IBS is characterised by abdominal pain and altered bowel habits (Thabane and Marshall, 2009). The odds of developing IBS are increased six-fold following an episode of acute GI infection, and increased risk may be associated with severe and prolonged GI infections, younger adults, and anxiety/depression (Thabane and Marshall, 2009; Thabane, Kottachchi and Marshall, 2007). Different studies estimate that between 0–18% and 0–38% of individuals with *Campylobacter* and non-typhoidal *Salmonella* infections, respectively go on to develop post-infectious IBS (Esan et al., 2017). It is thought that viral GI infections may give rise to a more transient form of post-infectious IBS, compared to bacterial GI infections (Marshall et al., 2007).
Guillain-Barré syndrome
Guillain-Barré syndrome is a rare autoimmune condition which affects the peripheral nervous system (WHO, 2016d). It is characterised by progressive weakness of the limbs which can be accompanied by numbness and/or pain (Yuki and Hartung, 2012). In severe cases the respiratory muscles are also affected which can be life-threatening (WHO, 2016d). Guillain-Barré syndrome is often preceded by an infection which can be bacterial or viral (WHO, 2016d). The most frequently identified infectious agent associated with Guillain-Barré syndrome is *Campylobacter jejuni* (Yuki and Hartung, 2012). Most studies report that <2% of individuals with *Campylobacter* infections will go on to develop the syndrome (Esan et al., 2017).
Haemolytic uraemic syndrome (HUS)
HUS is a rare condition, characterised by acute renal failure, haemolytic anaemia, and thrombocytopenia, and can be life-threatening (Kavanagh, Richards and Atkinson, 2008). It most commonly affects young children, and the vast majority of cases occur secondary to infection (Kavanagh, Richards and Atkinson, 2008; Tarr, Gordon and Chandler, 2005). HUS is most commonly associated with STEC infections, but other bacteria such as *Shigella dysenteriae* serotype 1, and certain viruses can induce HUS (Tarr, Gordon and Chandler, 2005). Around 10–15% of cases with an STEC infection will develop HUS (WHO, 2016b; Tarr, Gordon and Chandler, 2005).

Relatively little is known about the risk factors for poor outcomes from IID, especially in high income countries where poor outcomes are uncommon compared to low income

countries. Factors that compromise the immune system may increase vulnerability to severe illness (Lund and O'Brien, 2011). Infants have relatively immature immune systems (Simon, Hollander and McMichael, 2015), and thus are vulnerable to severe illness. Children are also at greater risk of severe dehydration compared to adults since water constitutes a greater proportion of a child's bodyweight (UNICEF and WHO, 2009). Additionally, young children have higher metabolic rates and thus use more water over the course of a day, and their kidneys are less able to conserve water compared to older children and adults (UNICEF and WHO, 2009).

A review of the literature which considered studies from both high and low income countries, found evidence to suggest that premature birth, low birth weight (1.5–2.49 kilograms), requirement of neonatal intensive care facilities, malnourishment and immunosuppression may be risk factors for severe rotavirus infections in children (Huppertz, Salman and Giaquinto, 2008). Breast feeding, on the other hand, has been shown to protect against prolonged diarrhoea in high and low income countries (Strand et al., 2012; Baker, Taylor and Henderson, 1998). Breast milk contains various antibodies (depending on prior maternal exposure), which provide protection against the GI infections likely to be encountered in the infant's environment (Turin and Ochoa, 2014). Additionally, breast milk directly contributes to the establishment of the infant's intestinal microbiota, which provides competition for nutrients and receptors thereby hindering the growth of pathogenic bacteria (Cacho and Lawrence, 2017). There is also evidence to suggest that breast milk can actively stimulate the infant's immune system, thus providing protection against GI infections for several years after the termination of breast feeding (Hanson et al., 2002).

At the opposite extreme of life, the immune system starts to deteriorate putting the elderly at increased risk of infection and poor outcomes following infection (Simon, Hollander and McMichael, 2015). As shown in Table 2.2, mortality due to IID in England and Wales is highest in the oldest age groups, and mortality starts to increase with advancing age from around 50 years of age. In addition to age, certain comorbidities and medications can predispose patients to more severe illness by decreasing their ability to combat infection or increasing the likelihood of organ failure (Farthing et al., 1996). In a review of the literature, Lund and O'Brien (2011) provide evidence to suggest that people with primary immunodeficiencies, transplant recipients, cancer patients, those with diseases of the immune system, human immunodeficiency virus (HIV) patients, people taking immunosuppressant drugs, the malnourished and those with cirrhosis or other liver disease, may be particularly vulnerable to foodborne infections and severe illness due to immune system compromise. Additionally, medications that lower stomach acidity, such as proton

pump inhibitors or H_2 receptor antagonists, also lower the body's defence against foodborne pathogens (Lund and O'Brien, 2011).

As detailed in Table 2.3, some pathogens, such as *Campylobacter* and *E. coli* O157, are associated with specific clinical complications of IID. On the other hand, prolonged diarrhoea can be caused by a number of different pathogens, and there is no clear evidence that any one particular pathogen is responsible for causing prolonged diarrhoea (Giannattasio, Guarino and Vecchio, 2016). In terms of mortality, a USA-based study analysed data from multiple surveillance systems and estimated that norovirus (58%), non-typhoidal *Salmonella* (11%), *Clostridium perfringens* (10%), and *Campylobacter* (9%) caused the most foodborne illness, but non-typhoidal *Salmonella* (28%), *Toxoplasma gondii* (24%), *Listeria monocytogenes* (19%), and norovirus (11%) were responsible for the majority of food-related deaths (Scallan et al., 2011). Another study conducted in the USA analysed mortality statistics and found that *Clostridium difficile* and norovirus were the two leading causes of gastroenteritis deaths across all age groups, although the elderly had the highest rates (Hall et al., 2012). Differences in the case definitions and study methodologies are likely to have contributed to the contrasting findings between these two studies. *Clostridium difficile* infections and norovirus outbreaks are common in hospital settings, and are associated with poor outcomes amongst already vulnerable populations (Fisher and Dembry, 2017; Iturriza-Gómara and Lopman, 2014).

A similar study to that performed by Scallan et al. (2011) was conducted in the UK, using surveillance data, outbreak data and data from the IID1 and IID2 studies. The results suggested that *Campylobacter*, *Clostridium perfringens* and norovirus cause the majority of foodborne illness, but the majority of hospitalisations are caused by *Salmonella* and *E. Coli* O157 (O'Brien et al., 2016). Differences between these estimates and the USA-based estimates may be due to the methods used, however *Salmonella* was identified as a leading cause of food-related mortality in the USA and hospitalisation in the UK. An Australian study estimated that *Salmonella* and *Campylobacter* caused the greatest disease burden (as measured by Disability Adjusted Life Years per case) in relation to morbidity, mortality and sequelae, compared to rotavirus, *Cryptosporidium*, *Giardia* and norovirus (Gibney et al., 2014). Other studies also suggest that *Toxoplasma gondii* and perinatal listeriosis cause high levels of individual disease burden, due to the sequelae and mortality associated with these pathogens, with perinatal listeriosis ranking highly due to the large number of years of life lost (Batz, Hoffmann and Morris, 2012; Havelaar et al., 2012; Lake et al., 2010). It is important to note however, that for the majority of cases presenting to healthcare services, a pathogen will not be identified.

A final point to mention is that the disease burden caused by any particular pathogen cannot be considered in isolation from host and environmental factors. In Austria, for example, hospitalised adults with diarrhoea caused by *Clostridium difficile* had higher all-cause mortality 30 days after discharge, compared to hospitalised adults with diarrhoea caused by pathogens other than *Clostridium difficile* (Schmid et al., 2014). This association was modified, however, by age and the presence of comorbidites. No statistically significant differences in the risk of mortality between the *Clostridium difficile* and non-*Clostridium difficile* patients were observed amongst those aged 65 years or older, those with additional infections or those with severe comorbidities. This study highlights the importance of host-related factors when considering the possible consequences and complications of GI infections.

Healthcare utilisation consequences

Healthcare utilisation as a consequence of GI infection will now be considered. There is evidence to suggest however that IID-related healthcare use is closely related to disease severity. The majority of individuals with IID do not seek healthcare advice but those that do tend to have more severe symptoms. This finding has been observed in countries such as the UK, France, the Netherlands, the USA and New Zealand (Doorduyn, Van Pelt and Havelaar, 2012; Van Cauteren et al., 2012; Adlam et al., 2011; Scallan et al., 2006; Tam, Rodrigues and O'Brien, 2003; De Wit et al., 2001b). In a Welsh survey, those with poorer self-rated health and comorbidites were more likely to consult a doctor for foodborne GI infections (Evans et al., 2006), and these factors may influence disease severity. There is also some evidence to suggest that doctor consultations for IID are more likely among young children and the elderly (age groups associated with vulnerability to severe illness), compared to younger adults (Van Cauteren et al., 2012; Adlam et al., 2011; Evans et al., 2006). Nonetheless, other studies have found that older adults are less likely to visit their GP for an IID compared to younger adults (Tam, Rodrigues and O'Brien, 2003).

As previously mentioned, estimates from the IID2 study indicate that around 2% of the UK population consult their GP for an episode of IID per year, equating to over one million consultations annually (Tam et al., 2012b). Additionally, in 2015–16 there were 154,020 hospital admissions with IID as the primary diagnosis as defined by International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) codes A00–A09 in England (Table 2.4). Of these, 110,483 were classified as emergency admissions, and

IID accounts for approximately 2% of all emergency hospital admissions in England (NHS Digital, 2016a). The vast majority of these admissions are of unknown aetiology (Table 2.4).

The burden on secondary care services is of particular importance, since most hospital admissions for gastroenteritis are considered to be preventable events. Gastroenteritis is classified as an ambulatory care-sensitive condition (ACSC), meaning that hospital admission for this condition could be avoided through early intervention and effective management (Ham, Imison and Jennings, 2010). In 2009–10 a report by think tank The King's Fund estimated that dehydration and gastroenteritis accounted for 10% of all ACSCs in England (Tian, Dixon and Gao, 2012). Other ACSCs include chronic conditions such as asthma, diabetes complications and chronic obstructive pulmonary disease; and acute conditions such as ear, nose and throat infections, dental conditions and epilepsy (Purdy et al., 2009). The King's Fund report also indicated that admission rates for ACSCs were largest for children aged <5 years, and the elderly aged >75 years. Additionally, admission rates in the most socioeconomically deprived areas were more than twice the rates in the most affluent areas (Tian, Dixon and Gao, 2012).

Socioeconomic disparities in hospital admissions for ACSCs have been observed in a number of countries, including those with universal access to healthcare such as the UK, and countries where healthcare is predominantly privately funded such as the USA (Kangovi et al., 2013). Reasons for these disparities may include increased prevalence of disease amongst more disadvantaged groups, or inadequate access to good quality primary care in deprived areas (Tian, Dixon and Gao, 2012).

Hospitalisation rates for ACSCs have been used as proxy measures for the quality of, and access to primary care services (Purdy et al., 2009; Ansari, 2007). Systematic reviews of chronic ACSCs suggest that improved access to primary care services, adequate primary care physician supply and increased continuity of care, may help to reduce hospitalisations for chronic ACSCs (van Loenen et al. 2014; Gibson, Segal and McDermott, 2013). A recently published UK-based study found that increased continuity of care was associated with reductions in hospitalisations for ACSCs, even after controlling for age, SES and level of comorbidity (Barker, Steventon and Deeny, 2017). In Spain, however, physician workload was not statistically significantly associated with rates of hospital admissions for ACSCs when socioeconomic variables were considered (Magán et al., 2011). These findings suggest that both SES and primary care quality may be important determinants of hospitalisations for ACSCs. Of note, some of the literature regarding ACSCs define gastroenteritis as purely non-infectious, and even within the NHS the diagnostic codes used to define gastroenteritis

as an ACSC vary considerably, with some sectors defining gastroenteritis as non-infectious (ICD-10 codes K52.2, K52.8 and K52.9) and others as infectious (ICD-10 codes A00–A09) (Purdy et al., 2009).

Table 2.4 Hospital admissions by primary diagnosis in England 2015–16

Primary diagnosis: ICD-10 codes A00–A09		Admissions	Emergency admissions
A00	Cholera	9	2
A01	Typhoid and paratyphoid fevers	228	193
A02	Other Salmonella infections	613	581
A03	Shigellosis	115	112
A04.0	Enteropathogenic Escherichia coli infection	3	3
A04.1	Enterotoxigenic Escherichia coli infection	4	1
A04.2	Enteroinvasive Escherichia coli infection	2	2
A04.3	Enterohaemorrhagic Escherichia coli infection	25	22
A04.4	Other intestinal Escherichia coli infections	144	134
A04.5	Campylobacter enteritis	2473	2420
A04.6	Enteritis due to Yersinia enterocolitica	3	3
A04.7	Enterocolitis due to Clostridium difficile	4562	4265
A04.8	Other specified bacterial intestinal infections	4037	269
A04.9	Bacterial intestinal infection, unspecified	134	115
A05.0	Foodborne staphylococcal intoxication	1	1
A05.1	Botulism	4	4
A05.3	Foodborne Vibrio parahaemolyticus intoxication	2	2
A05.4	Foodborne Bacillus cereus intoxication	3	3
A05.8	Other specified bacterial foodborne intoxications	6	6
A05.9	Bacterial foodborne intoxication, unspecified	160	159
A06	Amoebiasis	55	34
A07.1	Giardiasis [lambliasis]	170	88
A07.2	Cryptosporidiosis	147	139
A07.3	Isosporiasis	1	1
A07.8	Other specified protozoal intestinal diseases	5	4
A07.9	Protozoal intestinal disease, unspecified	1	1
A08.0	Rotaviral enteritis	573	553
A08.1	Acute gastroenteropathy due to Norwalk agent	622	581
A08.2	Adenoviral enteritis	124	118
A08.3	Other viral enteritis	972	907
A08.4	Viral intestinal infection, unspecified	30,462	30,048
A08.5	Other specified intestinal infections	34	29
A09	Other gastroenteritis and colitis of infectious and unspecified origin	108,326	69,683
	Total	**154,020**	**110,483**

Data source: NHS Digital (2016a)

Social and economic consequences

In 2010, the FSA estimated that foodborne illness alone cost the UK around £1.9 billion per annum (FSA, 2012). A proportion of these costs are borne by the NHS in terms of GP visits, hospitalisations, prescriptions and laboratory tests. Whilst hospital admission for IID tends to occur relatively infrequently, the associated costs can be substantial. Analysis of the IID1 study data suggested that hospital costs represent approximately 30% of the total NHS costs for managing IID (Roberts et al., 2003). In 2009–10, it was estimated that emergency hospital admissions for dehydration and gastroenteritis cost the NHS in England just under £128 million, although the definition of gastroenteritis was not specified (Tian, Dixon and Gao, 2012). As discussed, hospitalisation for gastroenteritis is considered to be a preventable event, and thus hospital admissions for this condition represent expensive, yet potentially avoidable costs.

Yet healthcare costs are not the only costs associated with IID. In the IID1 study, economic costs due to lost employment represented a significant proportion of the overall economic burden due to IID (Roberts et al., 2003). Economic costs due to lost earnings may affect the individual as well as society in terms of lost productivity. A UK-based study found that the average cost to the health sector for an episode of acute gastroenteritis in children aged less than five years was £60, but the total cost to society was £176 when accounting for parental costs and the value of work time lost (Lorgelly et al., 2008).

Furthermore, sickness absence is a common consequence of IID. Analysis of data collected in the IID2 study suggests that around 50% of individuals with IID report absence from work, school or daily activities due to their illness (Tam et al., 2012b). This represents approximately eight million absences from school, and more than 11 million days lost from employment amongst people of working age per year (FSA, 2016). In New Zealand, 90% of surveyed IID cases reported loss of time at work, school or recreation, and 36% reported missed work time (Adlam et al., 2011). Similar findings have been observed amongst adults surveyed in Germany and France with 23% and 21% of IID cases reporting work absenteeism, respectively (Wilking et al., 2013; Van Cauteren et al., 2012).

Summary

The consequences of GI infections have been discussed in three separate sections (disease severity/clinical, healthcare utilisation and social/economic) however it is evident that these consequences are related. Disease severity is likely associated with healthcare utilisation which in turn influences the social and economic costs of GI infections. Whilst GI infections are usually mild and self-limiting, occasionally severe clinical consequences can result. Furthermore, the high frequency with which GI infections occur in the community can amount to substantial overall societal costs.

The severity of GI infections may be measured by several inter-related variables, such as the extent of dehydration, the duration/frequency of symptoms, the need for hospitalisation or the development of complications or outcomes such as death. Little is known about risk factors for severe infections, however the age of the host appears to play an important role, with young children and the elderly most at risk of severe illness. No one pathogen has been consistently found to cause particularly severe disease in relation to others. This may be because disease severity is not only related to the infecting pathogen, but also to host factors (such as immunocompetence), and environmental factors (such as the dose to which the host is exposed), and these factors may differ across social groups. Host factors may be of greater clinical importance when assessing risk of disease progression, since a pathogen will not be identified for the majority of cases.

Having provided an overview of GI infections and the potential consequences of such infections, the discussion will now focus on socioeconomic inequalities in health.

2.4 SOCIOECONOMIC INEQUALITIES IN HEALTH

Broadly speaking, differences in the health status of certain individuals or groups can be described as health inequalities. In the UK, health inequalities commonly refer to the distribution of health by socioeconomic position (Smith, Bambra and Hill, 2016). Socioeconomic inequalities in health arise when systematic differences in health exist between groups which occupy unequal positions in society (Graham, 2007). Such inequalities are the focus of this discussion, which examines the importance of health inequalities, the various theories that have been put forward to explain them, and the ways in which they can be measured.

What are health inequalities and why are they important?

For more than a century, researchers have compared measures of mortality and morbidity across socioeconomic groups (McKee and Pommerleu, 2005). They have found that health inequalities within countries exist across the world, reflecting differences in the opportunities to achieve good health between rich and poor (Morgan, 2006). In general, life expectancies and the prevalence of most diseases display a social gradient whereby the poorest in society experience greater levels of illness and premature death than those further up the socioeconomic scale (Wilkinson and Marmot, 2003). For example, men living in the least deprived areas of England can expect to live almost a decade longer than men living in the most deprived areas (life expectancy gap of 9.3 years), and for women the life expectancy gap is 7.4 years (Office for National Statistics [ONS], 2018).

These health inequalities are important because health is important. Health is outlined as a fundamental human right by the United Nations (UN) and the WHO (UN General Assembly, 1948; WHO, 1948). People value their health highly, and since health is necessary to live and to function, inequalities in health also represent inequalities in the ability to function (Galama and van Kippersluis, 2013; Marmot, 2013). Given the importance of health, many consider it unjust that individuals further down the socioeconomic ladder should experience worse health and a decreased ability to function compared to those with increased levels of wealth and social influence; especially considering that the degree of socioeconomic stratification of most societies can be altered with social and political action. In the public health community, health inequalities that are socially produced are widely regarded as unnecessary, avoidable, unfair and unjust (Whitehead and Dahlgren, 2007).

Socioeconomic inequalities in health have been observed between countries and within countries, and across a range of diseases, both non-communicable and infectious in nature. As mentioned previously, mortality rates for diarrhoea are far higher in low compared to high income countries, as are mortality rates for other infectious diseases such as tuberculosis and HIV leading to acquired immunodeficiency syndrome (AIDS) (WHO, 2015). These health inequalities to some extent reflect different levels of poverty and adverse living conditions between low and high income countries (Farthing et al., 2012). Additionally, whilst non-communicable diseases were perceived to predominantly affect high income countries, evidence suggests that incidence rates for some non-communicable diseases, such as stroke, are increasing in low income countries and exceeding rates observed in high income countries (Sommer et al., 2015; Feigin et al., 2009).

Health inequalities also persist within countries, and have been observed within high income countries as well as low/middle income countries (Hosseinpoor et al., 2012). Yet, far more is known about health inequalities within high income countries which tend to have more comprehensive data collection systems, and have the resources available to invest in health research (Graham H, 2009).

The UK in particular has a rich history of heath inequalities research, partly due to the availability of administrative data on mortality, occupation and residential location (the latter has been used as a proxy measure for social class) (Smith, Bambra and Hill, 2016). In 1980, the Black Report, published by the Department of Health and Social Security, underlined widening socioeconomic inequalities in health despite the introduction of the NHS in 1948, and showed that mortality risk was approximately double amongst men in unskilled manual occupations compared to men in professional occupations (Gray, 1982; Black et al., 1980). The Black Report and a prospective cohort study of British civil servants named the Whitehall study, demonstrated that inequalities in mortality not only affect the top and bottom of the social scale, but that there exists a social gradient, whereby the risk of mortality increases progressively as social position decreases (Marmot, 2005; Marmot, Shipley and Rose, 1984; Black et al., 1980). The Black Report drew widespread public attention to health inequalities, and is now regarded as a seminal document in health inequalities research (Bartley and Blane, 2016; Smith, Bambra and Hill, 2016).

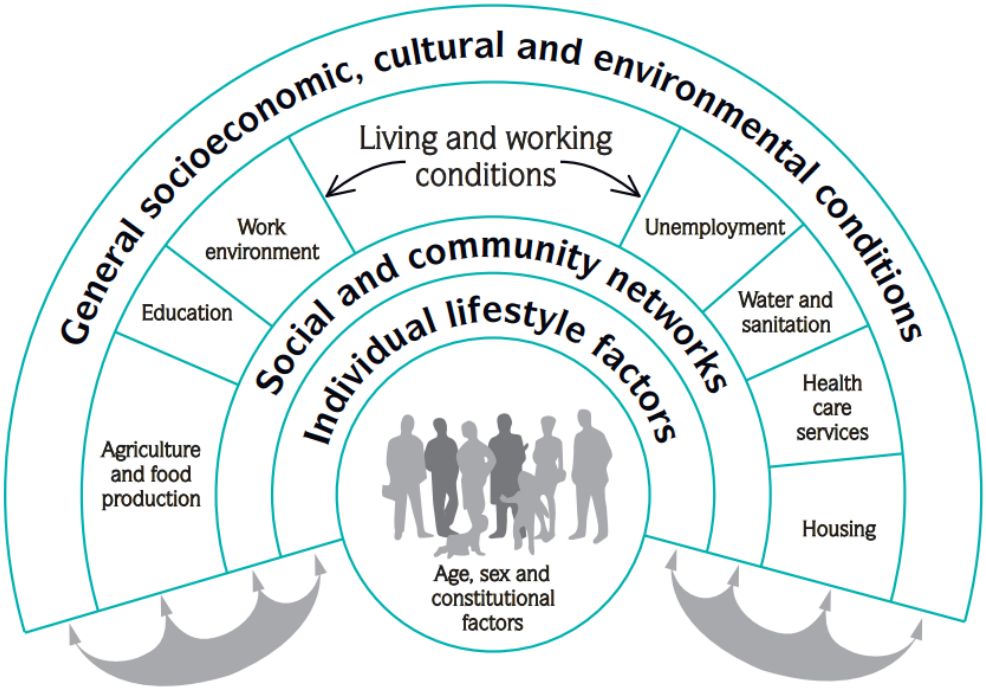
In the years following the Black Report, researchers have revealed that social gradients exist across a vast range of diseases, both in terms of incidence and outcomes including mortality. Examples of this include cardiovascular diseases, depression, suicide and other violent deaths, type II diabetes and other obesity related disorders, cancers, respiratory disorders, sexually transmitted diseases and neurological disorders (Crichton et al., 2015; Ramsay et al., 2014; Espelt et al., 2013; Siegrist and Marmot, 2006; Macleod, 2004; van Rossum et al., 2000). Health inequalities have also been shown to exist not only in terms of the burden of disease but also in the severity of disease and the number of years lived with disability and illness (Melzer et al., 2000; Whitehead, 1991). For example, men living in the least deprived areas of England compared to the most deprived can expect to live 15 years longer without a limiting longstanding illness or disability, and for women this gap is 13.5 years (ONS, 2013). As this evidence has accumulated, research has gradually shifted from describing health inequalities, to explaining and implementing interventions to address them (Mackenbach et al., 2002), as is subsequently discussed.

Determinants of health and health inequalities

Although it is widely recognised that heath inequalities exist, there is less consensus regarding explanations of why they exist. To begin, it might prove helpful to consider the factors that determine health.

Our health is determined by a wide range of factors spanning multiple ecological levels, as demonstrated in Figure 2.4. This model by Dahlgren and Whitehead (2007) illustrates the main determinants of health as a series of arcs that surround and exert their influence on individuals positioned centrally. The arrows emphasize the interactions that occur between the determinants, for example, individual lifestyles are embedded in social norms and living/working conditions, which in turn are related to the wider socioeconomic and cultural environment (Dahlgren and Whitehead, 2007).

Figure 2.4 Dahlgren and Whitehead's model of the determinants of health



Source: Dahlgren and Whitehead (2007)

These building blocks of health are often referred to as the social determinants of health, and are described by the WHO as the conditions in which people are born, grow, live, work and age, which are shaped by wider socioeconomic and political forces (Commission on Social Determinants of Health, 2008). The social determinants of health are considered to be the

most powerful determinants of health (over and above medical care), and are deemed to be responsible for the majority of the global burden of disease (Braveman and Gottlieb, 2014; Irwin and Scali, 2010; McGinnis, Williams-Russo and Knickman, 2002).

In terms of GI infections, living conditions relating to food quality, water and sanitation, and healthcare services are important social determinants of incidence and mortality rates globally. The macro level environment will dictate the policies that are in place to promote healthy living standards to prevent infections, and the level of support available to those who become ill. Social and community networks may also be important sources of support when recovering from an infection.

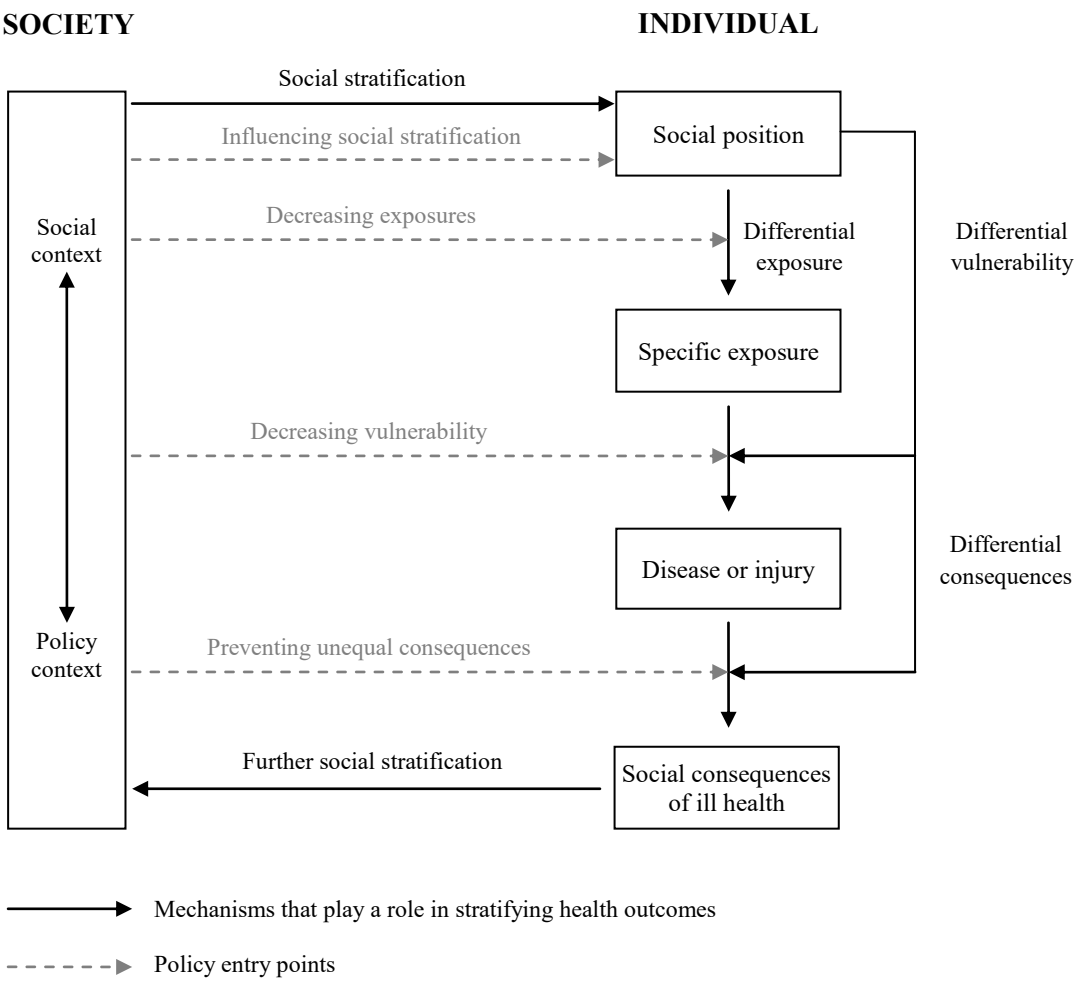
The social determinants of health might also be thought of as the social determinants of health inequalities, however Graham (2004) points out that it is necessary to distinguish between the social causes of health, and the factors and processes that determine their unequal distribution across the social hierarchy. In general, inequalities in health exist because both the access to resources which promote heath and the exposure to risks which damage health are determined by an individual's social position, and thus social position is a key determinant of health (Graham H, 2009; Graham, 2007; Link and Phelan, 1995). This concept is captured in Diderichsen's model of the mechanisms of health inequality (Figure 2.5), initially developed by Diderichsen and Hallqvist (1998) and later refined by Diderichsen, Evans and Whitehead (2001).

Diderichsen's model conceptualises how the social context of a society influences the social stratification of individuals within that society, and determines how equally or unequally resources, such as power and wealth, are shared amongst individuals of different social positions (Diderichsen, Evans and Whitehead, 2001). The model then outlines the potential mechanisms by which social contexts and positions might influence both health outcomes and the resulting consequences of ill health. Firstly, an individual's social position may determine their exposure to health damaging risks, and their vulnerability to ill health following such exposure. For example, individuals in lower compared to higher social positions are more likely to be exposed to various health damaging risks, and over time these might interact and eventually overwhelm their biological defences against such health damaging exposures (Diderichsen, Evans and Whitehead, 2001). Vulnerability to ill health not only relates to biological susceptibility but can also be linked to social factors such as lack of social support and psychosocial well-being (Diderichsen, Evans and Whitehead, 2001). Longitudinal studies have shown that health damaging exposures not only tend to cluster amongst the socioeconomically disadvantaged, but they also accumulate across the

life course (Blane, 2005). As a result, early life disadvantage can have a profound influence on adult health outcomes (Giesinger et al., 2014; Power and Kuh, 2006).

An individual's social position may also affect the consequences they experience following a disease event. The social consequences of ill health for the individual may include excessive healthcare costs and loss of work/income, which may have more damaging effects for those in lower social positions who have less financial cushioning, compared to those in higher positions (Diderichsen, Evans and Whitehead, 2001). Thus ill health may result in further social stratification, although this may be moderated to some extent by the social context, for example the provision and scope of welfare systems and universal healthcare coverage.

Figure 2.5 Diderichsen's model of the mechanisms of health inequality



Source: Adapted from Diderichsen, Evans and Whitehead (2001)

The model provides a theoretical framework of how social context and social position are linked to inequalities in health outcomes and the consequences of ill health, and it outlines the specific mechanisms involved, i.e. differential exposures, vulnerabilities and consequences. When considered together, the models that have been discussed (Figures 2.4 and 2.5) can enhance understanding of the social determinants of health, and the mechanisms involved in the unequal distribution of these determinants that ultimately lead to health inequalities. In addition, further insight may be gleaned by considering whether certain aspects of the social determinants of health play a greater role in generating health inequalities than others.

Exploring the various theoretical explanations of health inequalities may offer some clues as to whether certain exposures and vulnerabilities might contribute more towards the establishment of health inequalities than others. Several explanations of health inequalities have been put forward by researchers over time, and a selection of some of the most well known is subsequently discussed. These different explanatory models for health inequalities have, to some extent, a different emphasis on particular social determinants of health, i.e. lifestyle/behavioural factors, physical living and working conditions, psychosocial factors and the macro-level political and economic environment.

As mentioned previously, the Black Report drew attention to health inequalities in the 1970s, and within the report several theoretical explanations for the association between occupational class and health were put forward (Black et al., 1980). These are today regarded as 'traditional' explanations of health inequalities, yet they provided a theoretical framework and shaped the literature and discussion of health inequalities for decades (Siegrist and Marmot, 2006). One explanation provided in the Black Report, suggests that health-related behaviours, influenced by the cultural environment, are responsible for health inequalities. Such behaviours and lifestyle factors might include smoking, drinking alcohol and consuming high fat/sugar foods, which are all more prevalent in lower socioeconomic groups, and are also associated with major chronic diseases (Siegrist and Marmot, 2006). Some of these lifestyle factors may also influence vulnerability to GI infections. For example, cigarette smoke is known to have immune suppressing properties (Mehta, Nazzal and Sadikot, 2008; Sopori, 2002).

One version of this explanation proposes that individuals freely choose to engage in health-damaging behaviours, and that the social gradient in health is completely accounted for by differences in these behaviours across socioeconomic groups (MacIntyre, 1997). Implicit in the notion that individuals freely choose to engage in these behaviours, is that individuals

must therefore freely choose to be unhealthy; a position that blames the individual for their poor health. As a counter-argument to this, research has shown that inequalities in indicators of physical and mental health emerge early in childhood (Rougeaux et al., 2017; Rutherford et al., 2017), and clearly children are not to blame for these inequalities. An alternative version of this explanation asserts that individuals do not necessarily freely choose to engage in health-damaging behaviours, and that health-related behaviours contribute towards health inequalities rather than explaining them completely (MacIntyre, 1997). Pursuing this latter version, researchers have suggested that an individual's social position may determine and constrain the choices they can make, and that individuals in lower social positions may adopt unhealthy behaviours in order to cope with their circumstances and material disadvantage (Dahlgren and Whitehead, 2007; Asthana and Halliday, 2006).

This leads on to another explanation of health inequalities proposed within the Black Report, the materialist explanation. This explanation focuses on the direct influence of poverty and material deprivation in determining health outcomes, and suggests that an individual's level of income determines the conditions in which they live and work, and thus their exposure to physical health-damaging risks (e.g. poor housing, air pollution, occupational hazards) and access to health-promoting factors such as healthcare (Smith, Bambra and Hill, 2016; Siegrist and Marmot, 2006). This explanation was favoured within the Black Report, and many researchers agree that material deprivation and absolute poverty contribute to the observed inequalities in health (Graham H, 2009; Mackenbach et al., 2002; Whitehead, 1992; Black et al., 1980). However, it is also argued that this explanation alone cannot explain why health inequalities persist in high income countries with low levels of absolute poverty (Marmot, 2013; Mackenbach, 2012).

A suggestion of why health inequalities persist in high income countries with high living standards is that individuals in lower social positions are more exposed to psychological and social stresses which lead to poor health outcomes over time. It is proposed that psychosocial factors, such as feelings of social inferiority, social isolation, low levels of control and early life stress, are powerful determinants of health and health inequalities in affluent countries (Whitehead et al., 2016; Wilkinson, 2005). These psychosocial factors may influence negative health outcomes via the biological effects of chronic stress, or via the engagement of health-damaging behaviours as coping mechanisms (Kristenson, 2006; Steptoe, 2006; Brunner and Marmot, 2005). The importance of psychosocial factors in the workplace was outlined in the Whitehall II study, where it was observed that low levels of control at work accounted for much of the social gradient in coronary heart disease incidence (Marmot et al., 1997). Chronic stress has also been shown to have negative effects on immune system

functioning, suppressing the body's ability to initiate an efficient immune response to infection (Salleh, 2008; Segerstrom and Miller, 2004).

Psychosocial factors perhaps offer a better explanation for the stepwise social gradient in health outcomes, compared to material deprivation which is likely to affect the poorest in a society (Asthana and Halliday, 2006). Thus the psychosocial explanation of health inequalities addresses some of the gaps in the materialist explanation which can be criticised for overly focusing on physical living and working conditions (Siegrist and Marmot, 2006). Alternatively, rather than viewing materialist and psychosocial explanations separately, a better explanation may be that adverse material conditions and psychosocial factors interact to increase vulnerability to disease (Asthana and Halliday, 2006; Siegrist and Marmot, 2006; MacIntyre, 1997).

The psychosocial explanation of the social gradient in health asserts that in the most affluent countries of the world, relative income inequality is a more powerful driver of inequalities in health, than material living conditions alone (Wilkinson, 2005). Indeed, several studies have found that countries or populations with higher compared to lower levels of income inequality, also exhibit higher mortality rates, lower average life expectancies and lower levels of self-rated health (Wilkinson and Pickett, 2008; Subramanian and Kawachi, 2006; Kaplan et al., 1996; Wilkinson, 1992). Some argue that these findings reflect the socially corrosive nature of income inequality, which negatively affects rich and poor alike by increasing stress provoking social status insecurities and competition (Wilkinson and Pickett, 2008; Subramanian and Kawachi, 2006). Alternatively, others have proposed that the association between income inequality and population health can be explained entirely by political inequality (Deaton, 2013). In the USA, researchers have observed that government policy is uniquely responsive to the preferences of affluent individuals, likely due to their propensity to donate funds to parties, candidates and interest organisations (Gilens, 2005). The very rich do not benefit from social spending on health, education and welfare, and they may prefer to avoid paying taxes to support such spending (Deaton, 2013; Curtis, 2004). Thus, the concentration of wealth and the subsequent concentration of power in the hands of a few can have deleterious consequences for the majority, which might explain the association between income inequality and population health outcomes.

Reflecting again on the explanations of health inequalities expressed in the Black Report, an alternative theory to those discussed thus far is that of natural or social selection. This theory asserts that health determines social position, and this is captured to some extent in Diderichsen's model of the mechanisms of health inequality which suggests that, depending

on the social context, the economic costs of ill health may result in further social stratification. Similarly, poor health associated with child poverty may have a negative impact on children's cognitive, social-behavioural and educational outcomes, resulting in poor health and life chances in adulthood (Wickham et al., 2016). Some authors propose that the reverse causality effect is highly influential in explaining health inequalities, and that health has a strong effect on labour force participation, earnings and wealth (Galama and van Kippersluis, 2013). Conversely, researchers who have conducted longitudinal studies have found that the effects of health on social position are small, and have reached the conclusion that reverse causality is unlikely to be a primary explanation of the social gradient in health (Chandola et al., 2003; Hart, Davey Smith and Blane, 1998).

A final explanation is that the observed inequalities in health are an artefact of the way that the variables under study are measured (Black et al., 1980). Certainly there are many biases that can affect observational studies, and researchers have a duty to evaluate the potential biases and limitations within their work. However, it is widely regarded that statistical inaccuracies are unlikely to fully account for the magnitude and persistence of health inequalities over time (Graham H, 2009).

Reflecting on this discussion, there are a various explanatory models for health inequalities, and researchers tend to differ in their views as to the relative importance of these models in explaining health inequalities. However, a number of researchers agree that these explanatory models are not mutually exclusive, and that the underlying causes of health inequalities are likely to involve complex interactions between many social determinants of health (Smith, Bambra and Hill, 2016; Whitehead, 2007; Siegrist and Marmot, 2006). In terms of GI infections for example, psychosocial factors such as chronic stress, material factors such as malnutrition, and behavioural factors such as smoking may all contribute to inequalities by compromising immune functioning.

Measuring health inequalities

As previously discussed, social position is a key determinant of health, a concept which is captured within Diderichsen's model of the mechanisms of health inequality (Diderichsen, Evans and Whitehead, 2001). In health research, a person's socioeconomic position or SES is often assigned based on their level of education, occupation or income. These measures may capture the material and behavioural conditions of individuals which are associated with disease, but may fail to adequately capture the macro level political and societal forces which

generate social stratification (Graham, 2007; Mackenbach et al., 2002). Additionally, education, occupation and income may represent different aspects of socioeconomic position, however Siegrist and Marmot (2006) view this positively and argue that the different measures can be used to enhance understanding of causal mechanisms between socioeconomic position and disease.

For example, a person's level of education may determine how they obtain and interpret health related information, i.e. their health literacy, and how they interact with health services and healthcare professionals in order to maximise and improve their health (Deaton, 2013; Van der Heide et al., 2013). Education could therefore be viewed as having a direct effect on health, as well as indirect effects via job prospects and earning potential. Measuring SES by education level has certain advantages in that a person's level of education does not tend to vary over time, as can be the case with income or occupation. Conversely this can also be a drawback, since education level is a predictor of adult SES rather than a real-time measure of adult SES (Graham, 2007).

Income can be used as a measure of adult SES, and household income is often used to reflect the socioeconomic circumstances of children. Income determines access to material resources and may also reflect aspects of power and social standing (Graham, 2007; Siegrist and Marmot, 2006). Level of income can however vary across the life course, and falling ill may have a negative effect on a person's income.

Occupation is also often employed as a measure of adult SES, and has been used to represent social class in the UK since the 19th century (Graham H, 2009). A criticism of occupation as a measure of SES is that it can only be applied to employed individuals (Graham, 2007), however not all occupationally-based SES measures require individuals to be in current employment. For example, the National Statistics Socioeconomic Classification (NS-SEC) is derived based on the current or last main job of the main-earner in a household, and is a measure of employment relations and conditions of occupations (Rose, Pevalin and O'Reilly, 2005; ONS, no date). The NS-SEC is utilised in Study 2 of this thesis where individuals are classified according to the occupation of the main-earner in their household, with managerial/professional, intermediate and routine/manual occupations representing SES from high to low.

Individual measures, such as occupation, education and income, may each capture a distinct aspect of SES but usually correlate well with each other, and other individually-based measures of SES (Shavers, 2007). Individual-level data can be aggregated to form area-level

measures, however area-level measures of SES sometimes do not correlate well with individual SES (Pardo-Crespo et al., 2013; Shavers, 2007). Aggregation can be performed over large geographical areas such as Local Authorities in the UK, or small areas such as Lower-Layer Super Output Areas (LSOAs) which were introduced following the 2001 Census and each contain approximately 1000 to 3000 people (ONS, 2016). The English Index of Multiple Deprivation (IMD), for example, is a relative measure of deprivation provided at the LSOA level, and consists of multiple domains of deprivation relating to income, employment, education, health, crime, the living environment and barriers to housing and services (Department for Communities and Local Government, 2015).

Area-level measures of SES may capture the socioeconomic characteristics of the individual residents of an area, but these measures may also capture characteristics of the area itself. Research has shown that these neighbourhood characteristics, otherwise known as contextual factors, can affect health behaviours and health outcomes independently of individual SES (Bambra, 2016; MacIntyre and Ellaway, 2009; Sellström and Bremberg, 2006; Pickett and Pearl, 2001). It would appear that regardless of individual SES, living in an economically deprived neighbourhood confers additional increased risk of premature mortality and long term illness (Ross and Mirowsky, 2008; Van Lenthe, 2006; Pickett and Pearl, 2001).

There are various theories of how the contextual socioeconomic environment influences health outcomes. The socioeconomic context of a neighbourhood may have a direct affect on health, or may influence health indirectly via mechanisms such as the availability of services, the physical environment, community cohesion, and the stigma or reputation associated with an area (Bambra, 2016; MacIntyre and Ellaway, 2009; Pickett and Pearl, 2001). These neighbourhood characteristics may be related. For example, a neighbourhood could have a negative reputation due to some aspect of the physical environment such as a toxic waste dump, and both the physical exposure and psychosocial stress associated with living in such an area may increase residents' risk of ill health (Bambra, 2016).

In terms of GI infections, there are several contextual factors which could be related to the risks and consequences of infection. For example, a person's risk of acquiring a GI infection may be associated with the quality of the local food environment. A study by Collins (2013) found that in England, food establishments with the lowest hygiene ratings were more concentrated in the most deprived neighbourhoods. Furthermore, individual behaviours, such as breastfeeding which is known to reduce GI infection risk in infants (Stuebe, 2009), may be strongly influenced by community and cultural norms (Swanson et al., 2017). Consequences of infection such as sickness absence may be associated with the level of

social cohesion in a neighbourhood, with close-knit communities providing support and care for those who are unwell. On the other hand, GI infections may be transmitted more easily in communities with strong social networks, where individuals visit each other and meet frequently. Additionally, variation between neighbourhoods in the quality of healthcare services and the ease with which they can be accessed may be important determinants of the consequences of GI infections.

The impact of neighbourhood characteristics on health may also vary by individual-level factors such as age, ethnicity and sex. For example, a UK-based study found that neighbourhood characteristics such as low levels of trust and tolerance, reduced access to banks and health services, and low quality physical environments were associated with poorer self-rated health for women but not for men when individual SES was accounted for (Stafford et al., 2005). For men, individual SES, age and family type explained all of the between-neighbourhood variation in self-rated health (Stafford et al., 2005). These findings indicate that in the UK, contextual socioeconomic characteristics may have a greater impact on women's self-rated health compared to men's.

The importance of the contextual socioeconomic environment on health also tends to vary by country (MacIntyre and Ellaway, 2009). Possible explanations for this include differences in levels of inequality, differences in residential segregation by individual SES, and differences in levels of exposure to neighbourhood-level factors between countries (Van Lenthe, 2006). This variation between countries highlights the importance of considering the role that wider political and economic forces play in determining health alongside individual and contextual factors (Bambra, 2016). The degree of social stratification and segregation in a society is determined by the political economy, and the reasons why deprived neighbourhoods and poverty exist are because the political economy allows them to exist (Bambra, 2016).

Thus, area-level measures of SES may capture both the socioeconomic characteristics of individual residents as well as contextual factors shared by communities, however similar to individual-level measures, they may inadequately capture the wider political and societal forces which generate social stratification. Another limitation of area-level measures is that characteristics of the aggregated population can be incorrectly attributed to individuals within the population (Curtis, 2004). Ecological bias can occur when associations present at the area-level are not apparent at the individual-level, possibly due to measurement error or confounding (Greenland and Robins, 1994). Ecological bias may be especially problematic when data on individual SES are aggregated over large populations, and selecting smaller areas over which to aggregate data may reduce the risk of measurement error. Having

provided some background information on both GI infections and health inequalities, the remainder of this chapter seeks to draw these two topics together.

2.5 SOCIOECONOMIC INEQUALITIES IN GI INFECTIONS

In this section, I review the literature and current knowledge relating to socioeconomic inequalities in GI infections in high income countries. Particular focus is paid to studies that have investigated inequalities in healthcare presentation for GI infections, and in the incidence of infection in the community, since there are a number of studies on these topics. I highlight differences and similarities between these studies, and identify several gaps in the knowledge base. The three studies presented in this thesis attempt to address some of these gaps, in order to enhance current understanding of inequalities in the consequences of GI infections.

Inequalities in healthcare presentation for GI infections

The starting point of this thesis rests on observations made by some studies that inequalities in healthcare presentation for GI infections are apparent. Several studies have sampled primary or secondary care presenting GI infection cases, and compared the socioeconomic distribution of these cases to the socioeconomic distribution of individuals in the general population. A smaller group of studies have investigated inequalities in healthcare presentation for GI infections, amongst diagnosed GI infection cases only. This latter group of studies may reveal more information about healthcare utilisation as a consequence of GI infection, since their interpretation is not hindered by uncertainties in the social patterning of the incidence of infection in the general population.

The studies that have been mentioned are noticeably heterogeneous in terms of study design factors. Various case definitions and measures of SES have been used. For example, individual-level measures of SES such as occupation and education level, as well as area-level measures such as the IMD. The studies have also been conducted in a number of different countries with varying levels of economic development and healthcare provision. These studies are subsequently discussed, using illustrative examples to highlight the differences and similarities between them.

Firstly, amongst studies that have investigated differences in SES between GP presenting IID cases and general population comparison groups, some have found evidence to suggest that those of lower SES are at an increased risk of IID identified via GP presentation (Phillips et al., 2011; Beale et al., 2010; Teschke et al., 2010; Quigley et al., 2006). Although not all studies have observed this association (Arena et al., 2014; Rodrigues et al., 2001; Sethi et al., 2001). These studies have analysed individuals of different ages, and defined IID in various ways.

For example, inconsistent findings have been observed by four studies that analysed GP presenting IID cases, and age and sex matched population-based controls (registered at the same practice), identified via the IID1 study in England. Quigley et al. (2006) found that infants aged ≤1 year from lower social class households (based on occupation of the main wage-earner), had an increased risk of IID identified via GP presentation (odds ratio [OR] 2.14; 95% CI 1.19–3.85). Phillips et al. (2011) similarly found an increased risk of norovirus infection identified among children aged <5 years from lower social class households compared to high (OR 2.3; 95% CI 1.4–3.9). On the other hand, Sethi et al. (2001) found that children aged <16 years from lower social class households had a non-statistically significant increased risk of rotavirus infection identified via GP presentation (OR 1.2; 95% CI 0.71–2.04). Additionally, Rodrigues et al. (2001) found no statistically significant association between employment type (OR 0.9; 95% CI 0.5–1.6 for unemployed versus full-time employed) of the main wage-earner in the household, and risk of *Campylobacter* infection in individuals aged >1 year. Whilst each of these four studies analysed data from the English IID1 study, they investigated different pathogens, and only Rodrigues et al. (2001) analysed both children and adults combined.

One French case-control study by Arena et al. (2014), analysed adults (aged ≥18 years) only, and found that adults with high school education and above had a non-statistically significant increased risk of viral gastroenteritis identified via GP presentation compared to those with middle school education (OR 2.37; 95% CI 0.86–6.57) adjusting for public transport use, contact with gastroenteritis cases, children aged ≤2 years in the household, and professional status (employed or student, versus non-employed or retired). Controls were selected from the same GP as the cases arose, and were age and sex matched to the cases. The results of this study suggest that adults of lower SES compared to high are less likely to present to their GP with an IID; a finding which contrasts with the results of some of the studies mentioned above which analysed children only (Phillips et al., 2011; Quigley et al., 2006).

Additionally, amongst studies that have assessed social gradients in hospital admissions for IID, some have found an inverse relationship between SES and IID-related admission rates (Biering-Sørensen et al., 2012; Lal et al., 2012; Wilking et al., 2012; Pockett et al., 2011; Moorin et al., 2010; Ma, El Khoury and Itzler, 2009; Özmert, Kilic and Yurdakök, 2008; Dennehy et al., 2006; Olowokure et al., 1999; Borgnolo et al., 1996), whilst others have not observed this association (Xu, Hu and Tong, 2015; Seo et al., 2013; Kyle et al., 2011; Teschke et al., 2010; Kum-Nji et al., 2009). These studies have been conducted in a number of countries across the world including the UK, Denmark, Germany and the USA. Most have used ecological study designs, but cohort and case-control designs have also been employed.

In the UK, three ecological studies have examined the relationship between SES and IID-related hospital admissions, with somewhat inconsistent findings. Of note, none of these studies controlled for potential confounding variables in their analyses, however one study (Olowokure et al., 1999) stratified results by age group. Pockett et al. (2011) found a statistically significant association between increasing hospital admissions rates for children aged <5 years with infectious gastroenteritis, and increasing area-level deprivation (measured using the IMD) across England. On the other hand, Kyle et al. (2011) analysed data from the Greater London region, and found no statistically significant correlations between the IMD and emergency hospitalisations for diarrhoea in children aged 0–14 years. Lastly, Olowokure et al. (1999) analysed data from the West Midlands, and found that admission rates for IID increased with increasing deprivation across five age groups (ranging from 0 to >75 years). Children aged 0–4 years, and the elderly aged ≥75 years had the largest differences in admission rates between the most and least deprived quintiles. This study is among the few that have investigated the relationship between SES and risk of IID-related hospitalisation for adults specifically.

Elsewhere in Europe, studies have utilised various study designs, but have observed similar findings. For example, Biering-Sørensen et al. (2012) analysed individual-level data from a prospective Danish cohort study and found that amongst children aged <6 years, those of parents with low educational status (10^{th} grade) had a greater risk of being hospitalised for IID, compared to children of parents with a master's degree or higher education, adjusting for household income and labour market attachment (for maternal education hazard ratio [HR] 1.52, 95% CI 1.44–1.61; for paternal education HR 1.18, 95% CI 1.13–1.24). Additionally, Wilking et al. (2012) performed an ecological analysis in Berlin, Germany, and found that the incidence of laboratory-confirmed rotavirus hospitalisations among children aged <4 years, increased by 4.95% for each percent increase in unemployed inhabitants in the neighbourhood, when controlling for migration volume, population density, proportion of

foreign residents, residential quality and day care attendance. Similar associations were observed among adults aged ≥60 years in univariate and multivariate analyses, however these did not reach statistical significance.

The USA is a country which does not have universal healthcare coverage and healthcare is predominantly privately funded via private insurance plans (National Institutes of Health, 2016; Tanne, 2007). However, there is some evidence of inequalities in secondary care use for IID amongst American children, similar to that observed in other countries with more universal healthcare systems such as the UK. The studies conducted in the USA have all analysed children. Ma, El Khoury and Itzler (2009) conducted an ecological study and found rates of hospitalisations for gastroenteritis were higher amongst children aged <5 years enrolled in Medicaid (101.2 per 10,000 children), compared to children not enrolled (such as those with private health insurance) (64.3 per 10,000 children). Similar findings were observed for rotavirus hospitalisations. Medicaid is a government funded health insurance program for certain individuals and families with low incomes, however poverty alone does not necessary qualify individuals or families for Medicaid (Stephens, 2012). In a case-control study, Dennehy et al. (2006) found children aged ≤59 months of mothers with less than high school graduate education had a marginally significant increased risk of being hospitalised for rotavirus gastroenteritis (OR 1.5; 95% CI 1.0–2.3) compared to children of mothers with at least high school graduate education, adjusting for breast feeding, child care, children in the household, maternal age and child's age, sex, ethnicity and birth weight. In contrast, Kum-Nji et al. (2009) analysed data from a prospective cohort study, and did not observe any statistically significant associations between maternal employment status (employed versus unemployed relative risk [RR] 1.19; 95% CI 0.65–2.16) or health insurance (public versus private RR 1.30; 95% CI 0.48–3.54) and acute gastroenteritis amongst children aged ≤36 months, identified via clinic visits, emergency department visits or hospitalisations.

In contrast to the majority of studies mentioned thus far, the only study that found a statistically significant lower risk of secondary healthcare use for individuals of lower SES was a case-control study conducted in South Korea. Seo et al. (2013) analysed age and sex matched cases and controls of all ages, and found a statistically significant lower risk of secondary care presentation for hepatitis A infection amongst individuals of lower SES. The univariate odds ratio for individuals with middle school education or lower, compared to college education or higher was 0.39 (95% CI 0.17–0.93).

A final group of studies to be considered are those that have analysed GI infection cases only (Doorduyn, Van Pelt and Havelaar, 2012; Van Cauteren et al., 2012; Scallan et al., 2006; Tam, Rodrigues and O'Brien, 2003; Herikstad et al., 2002; De Wit et al., 2001b). Again, some of these studies have found evidence to suggest those of lower SES compared to high are more likely to present to healthcare services with an IID. Additionally, the studies suggest that IID cases with more severe symptoms are more likely to present to primary care services. This observation was consistently observed amongst the studies that measured disease severity using a symptom-based severity score, or the duration of symptoms, as detailed below.

Tam, Rodrigues and O'Brien (2003) analysed IID cases aged ≥16 years from the English IID1 study to investigate risk factors for GP presentation for IID. In multivariate analysis, the authors found that amongst IID cases, low education level was associated with greater odds of GP presentation for IID. Individuals with IID who left full-time education before 16 years of age, had twice the odds of presenting to their GP for their illness, compared to those who left at 19 years of age or older (OR 2.06; 95% CI 1.22–3.50). The authors also found that the strongest predictor of GP presentation in their multivariate model was disease severity. Cases with severe illness had 12 times the odds of presenting to their GP compared to those with mild illness (OR 12.54; 95% CI 7.58–20.74). The sample size was however insufficient to assess the relationship between SES and IID severity.

Two similar studies were performed in the Netherlands. De Wit et al. (2001b) analysed self-reported gastroenteritis cases of all ages identified via a community-based prospective cohort study (Sensor study). There were no statistically significant differences in the odds of presenting to a GP among cases with a low or high education level, compared to cases with an intermediate education level in univariate analysis. Parental education level was used for child participants. The authors did however find that cases with higher symptom severity scores had greater odds of presenting to a GP in univariate analysis (OR 3.3; 95% CI 1.4–8.0). Additionally, Doorduyn, Van Pelt and Havelaar (2012) analysed data from a cross-sectional postal survey of individuals of all ages. Low education level (or low parental education level for children) was not related to the risk of acquiring an IID in this study (OR 1.1; 95% CI 0.7–1.9). However, amongst the IID cases only, those with low education level compared to intermediate, had six times the odds of visiting a physician due to their illness (OR 6.1; 95% CI 1.3–27.6). This association was attenuated and became non-significant after adjustment for duration of symptoms in multivariate analysis. Cases who vomited for three or more days compared to 1–2 days were more likely to visit a physician (OR 7.9; 95% CI 1.4–44.5). The confidence intervals for the estimates in this analysis were wide, however

the results are interesting since inequalities were observed in healthcare utilisation as a consequence of infection, but not in the risk of acquiring a symptomatic infection.

Similar findings were observed by two USA-based studies which analysed data collected in telephone-based population surveys. Herikstad et al. (2002) analysed individuals of all ages who were contacted via random digit dialling. The percentage of participants who reported experiencing diarrhoea in the four weeks prior to interview increased as education level (or parental education level for child participants) increased from less than high school level, to college graduate level. However, the reverse trend was seen for the percentage of cases with diarrhoea who visited a medical practitioner (16% of cases with less than high school education and 9% of college graduates cases), although this trend did not reach statistical significance. In a separate telephone-based survey, Scallan et al. (2006) found that amongst self-reported IID cases of all ages, those with annual household incomes of <$25,000 were more likely to seek medical care for their illness compared to cases with higher household incomes, even after controlling for age, sex, health insurance, illness duration and various symptoms (OR 1.89; 95% CI 1.31–2.72). Cases who experienced diarrhoea for three or more days were also more likely to seek medical care.

Finally, in a French telephone-based population survey of individuals of all ages, Van Cauteren et al. (2012) found that the risk of self-reported acute gastroenteritis was lower amongst individuals whose head of household had a lower education level. However, amongst the IID cases only, the education level of the head of household was not statistically significantly associated with healthcare presentation for acute gastroenteritis. Individuals with longer duration of illness (>3 days versus ≤3 days) were statistically significantly more likely to have consulted a doctor for their illness in multivariate analysis controlling for age, sex and presence of headache (OR 4.55; 95% CI 2.16–9.59).

Inequalities in the incidence of GI infection in the community

Whilst there are a number of population-based surveys that have investigated the association between SES and the risk of acquiring a GI infection in the community, the direction of this association remains unclear, and conflicting findings have been observed by several studies. Some have found an increased risk of GI infection amongst those of lower SES (Beale et al., 2010; Özkan et al., 2007; Ludvigsson et al., 2006; Etiler, Velipasaoglu and Aktekin, 2004; Bozkurt, Özgür and Özçirpici, 2003; Bozkurt, Özgür and Özçirpici, 1999; Baker, Taylor and Henderson, 1998; Turkish Ministry of Health, 1995; Eaton-Evans and Dugdale, 1987).

Whilst others have found a reduced risk of infection amongst those of lower SES (Adams et al., 2017; Pollard et al., 2014; Van Cauteren et al., 2012; Scallan et al., 2004; Herikstad et al., 2002; De Wit et al., 2001a; Fein, Lin and Levy, 1995). Numerous others have found no statistically significant association between SES and GI infection risk. These studies vary considerably in terms of study design factors and the populations studied, and as such there could be a number of explanations for the contrasting findings observed. This is illustrated by the following examples.

In the UK, Beale et al. (2010) performed a cross-sectional analysis using population-based survey data of mothers and infants aged <18 months, and used household council tax valuation bands from A to E+ as a measure of SES. Mothers in council tax band A (indicating lower SES) reported infant diarrhoea more frequently, in univariate analysis. In Sweden, Ludvigsson et al., (2006) performed a prospective cohort study, and found that low maternal education was associated with increased incidence of diarrhoea in children aged 2.5 years, in univariate analysis. This association did not remain statistically significant following adjustment for maternal smoking during pregnancy, infant sex, and accommodation type.

In Turkey, Etiler, Velipasaoglu and Aktekin (2004) and Bozkurt, Özgür and Özçirpici (1999) analysed data from a prospective cohort study and cross-sectional survey, respectively, and found higher rates of diarrhoea amongst young children (aged <5 years) whose parents received fewer years of school education. Neither of these studies controlled for confounding variables in their analyses. Another study conducted in Turkey by Özkan et al. (2007) found that households with lower monthly incomes had statistically significantly higher rates of diarrhoea in a cross-sectional survey of household members of all ages. This association was statistically adjusted for the occurrence of water shortages, and the distance between household well and septic tank.

Amongst the studies that have observed a lower risk of infection amongst individuals of lower SES, is a recent study by Adams et al. (2017), which was conducted as part of a PhD project examining inequalities in exposures and vulnerabilities to GI infections in the UK. Adams et al. (2017) performed a longitudinal analysis of data from the UK-based IID2 prospective cohort study, and found that IID risk was statistically significantly lower among those in routine/manual compared to managerial/professional occupations, controlling for age, sex, rurality and employment status. Individuals of all ages were analysed, and participants reported the occupation of the main wage earner in their household. A similar study conducted by De Wit et al. (2001a) analysed data from a prospective cohort study in

the Netherlands and found that amongst individuals of all ages, the incidence of acute gastroenteritis statistically significantly decreased as education level decreased, controlling for age, sex and previous history of diarrhoea. Parental education level was used for child participants.

Similar findings have been observed by telephone-based surveys, conducted in the UK, Europe, the USA and Australia. Scallan et al. (2004) conducted a telephone survey in Northern Ireland and the Republic of Ireland, and found that amongst individuals of all ages, those whose head of household had a manual compared to professional/non-manual occupation had a lower risk of self-reported acute gastroenteritis in multivariate analysis, adjusting for age, sex, region of residence, household size and household residents aged <18 years. In the USA, using similar telephone survey methodologies, Herikstad et al. (2002) and Fein, Lin and Levy (1995) found that individuals with lower levels of education had a lower risk of self-reported diarrhoea. Herikstad et al. (2002) analysed individuals of all ages and performed univariate analysis, whereas Fein, Lin and Levy (1995) analysed adults only, and performed multivariate analysis controlling for age, sex and ethnicity. Lastly, in Australia, Pollard et al. (2014) found that adults educated to university degree level had a statistically significantly higher risk of self-reported food-poisoning, controlling for age, residential area and the number of meals eaten away from home.

Discussion

In high income countries, there are several studies that have sampled primary or secondary care presenting GI infection cases, and compared the socioeconomic distribution of these cases to the socioeconomic distribution of individuals in the general population. Some of these studies suggest those of lower SES compared to high have higher rates of GP consultation and hospital admission due to GI infections (Biering-Sørensen et al., 2012; Lal et al., 2012; Wilking et al., 2012; Phillips et al., 2011; Pockett et al., 2011; Beale et al., 2010; Moorin et al., 2010; Ma, El Khoury and Itzler, 2009; Özmert, Kilic and Yurdakök, 2008; Dennehy et al., 2006; Quigley et al., 2006; Olowokure et al., 1999; Borgnolo et al., 1996). Although, not all studies have observed this association (Xu, Hu and Tong, 2015; Arena et al., 2014; Seo et al., 2013; Kyle et al., 2011; Teschke et al., 2010; Kum-Nji et al., 2009; Rodrigues et al., 2001; Sethi et al., 2001).

These studies have been conducted in a number of different countries, and have used various methodologies and case definitions. For example, some have analysed specific pathogens

(e.g. Wilking et al., 2012; Phillips et al., 2011; Rodrigues et al., 2001), whilst others have investigated IID of any cause (e.g. Quigley et al., 2006; Olowokure et al., 1999). However, a similarity amongst these studies is that the majority have analysed either paediatric populations only, or individuals of all ages combined. Based on previous evidence presented within this chapter, it seems reasonable to assume that both risk of infection and disease course/progression vary by age. For example, children are at a greater risk of severe dehydration compared to adults, since water constitutes a greater proportion of their bodyweight (UNICEF and WHO, 2009). Therefore it is likely that healthcare presentation rates for GI infections vary by age, and inequalities in healthcare presentation rates may also vary by age. Stratifying analyses by child and adult age groups may reveal useful insights that enhance understanding of inequalities in healthcare presentation for GI infections for adults and children.

A key issue in relation to the interpretation of the studies that have investigated inequalities in healthcare presentation rates for GI infections, is that it is difficult to know whether their results reflect differential incidence of infection, or differential healthcare utilisation by SES. A recent UK-based literature review found evidence to suggest that in general, those of lower SES compared to high tend to use more healthcare at any given age, because they are sicker (Cookson et al., 2016). If individuals of lower SES have a greater risk of symptomatic GI infections, it stands to reason that they would also have a greater need for healthcare services and thus be more likely to present. Thus, increased healthcare presentation for GI infections amongst those of lower SES, may simply reflect increased need due to higher incidence of infection in disadvantaged groups.

Reviewing the findings of studies that have measured inequalities in the risk of acquiring a GI infection in the community may therefore assist in the interpretation of studies that have investigated inequalities in healthcare presentation rates for GI infections. However, whilst there are several studies on this topic, the direction of the association between SES and GI infection risk remains unclear. A number of population-based studies have observed an increased risk of GI infection amongst more socioeconomically disadvantaged groups (Beale et al., 2010; Özkan et al., 2007; Ludvigsson et al., 2006; Etiler, Velipasaoglu and Aktekin, 2004; Bozkurt, Özgür and Özçirpici, 2003; Bozkurt, Özgür and Özçirpici, 1999; Baker, Taylor and Henderson, 1998; Turkish Ministry of Health, 1995; Eaton-Evans and Dugdale, 1987). Whilst others have observed the opposite; an increased risk of infection amongst more socioeconomically advantaged groups (Adams et al., 2017; Pollard et al., 2014; Van Cauteren et al., 2012; Scallan et al., 2004; Herikstad et al., 2002; De Wit et al., 2001a; Fein, Lin and Levy, 1995). There could be a number of potential explanations for the contrasting

findings observed. For example, the studies vary in terms of study design, case definitions, the populations studied and the measures of SES used. The studies also controlled for a number of different potential confounding variables, such as age, ethnicity, maternal smoking and household size. Thus, there are many sources of clinical and methodological heterogeneity amongst studies that have measured the association between SES and the risk of GI infections, and it is unknown to what extent these sources contribute to the contrasting results observed. Gaining a better understanding of these studies and inequalities in the risk of acquiring a GI infection, may also reveal insights as to whether inequalities in healthcare presentation rates for GI infections are proportionate to need. One systematic review has previously investigated the association between SES and the risk of laboratory confirmed foodborne GI infections in high income countries. Newman et al. (2015) identified 16 studies, and using a narrative synthesis concluded that the association between SES and foodborne GI infection risk differed by pathogen. A key limitation of this review, however, was the narrow focus on laboratory-based studies. Cases of GI infection identified via laboratories represent a small fraction of cases occurring in the community (Tam et al., 2012b) and these cases may differ in terms of SES, to community cases.

A separate group of studies to those that have been considered are those that have investigated inequalities in healthcare presentation for GI infections amongst GI infection cases only. These particular studies may reveal more information about healthcare utilisation as a consequence of GI infection, since their interpretation is not hindered by uncertainties in the social patterning of the incidence of infection in the general population. Amongst these studies, Doorduyn, Van Pelt and Havelaar (2012), Van Cauteren et al. (2012) and Herikstad et al. (2002) investigated inequalities in both the risk of infection in the community, and in the risk of healthcare presentation amongst cases only. Two of these studies (Doorduyn, Van Pelt and Havelaar, 2012; Herikstad et al., 2002) found some evidence that GI infection cases of lower SES were more likely to present to primary care; yet found either a lack of a relationship between SES and risk of infection in the community, or an increasing risk of infection with increasing SES. These findings suggest that rather than differential risk of infection by SES, alternative factors could help explain the social gradient in healthcare presentation rates observed by some studies.

One such potential explanation could be that once infected those of lower SES compared to high, experience more severe symptoms and are thus more likely to present to healthcare services. Almost all of the studies that analysed GI infection cases only, investigated whether disease severity was associated with primary care presentation. These studies measured disease severity using severity scores or the duration of symptoms, and all observed an

increased risk of GP presentation amongst cases with more severe infections (Doorduyn, Van Pelt and Havelaar, 2012; Van Cauteren et al., 2012; Scallan et al., 2006; Tam, Rodrigues and O'Brien, 2003; De Wit et al., 2001b). Additionally, some other studies have found evidence of a social gradient in the duration of IID, however these are sparse in number, and have exclusively studied paediatric populations (Ma, El Khoury and Itzler, 2009; Baker, Taylor and Henderson, 1998; Conway, Phillips and Panday, 1990).

Sickness absence might also be thought of as a measure of disease severity, however very few studies that have investigated the relationship between SES and sickness absence due to IID. Using a large cohort of UK civil servants, Feeney et al. (1998) observed that age adjusted rates of sickness absence due to gastroenteritis lasting seven days or less, were over six and four times higher for men and women respectively, in lower employment grades compared to high. A social gradient was also observed for absences lasting longer than seven days (Feeney et al, 1998). Conversely, self-reported sickness absence for gastroenteritis in a cohort of Dutch employees was unrelated to education level in univariate analysis (Mohren et al., 2005). Thus, the few studies that have investigated the relationship between SES and sickness absence due to IID have produced inconsistent results, and neither controlled for potential confounding variables such as symptom severity.

Since episodes of IID are usually self-limiting, the small percentage of those with an IID who present to secondary care services and are admitted to hospital, may represent a specific subset of cases with particularly severe symptoms. Thus, hospitalisation for IID could potentially be thought of as a measure of disease severity. Out of the three ecological studies that investigated inequalities in IID-related hospital admissions in the UK, two found evidence of a socio-spatial gradient in admission rates (Pockett et al., 2011; Olowokure et al., 1999), and one found no statistically significant relationship between area-level deprivation and admission rates (Kyle et al., 2011). Two of the studies aggregated data over large geographical areas (Primary Care Trusts [PCTs]) (Pockett et al., 2011; Kyle et al., 2011), and none controlled for potential confounding variables in their analyses other than age. Olowokure et al. (1999) did stratify their results by child and adult age groups, however the data they analysed were collected over 20 years ago (between 1990 and 1995).

2.6 GAPS IN THE EVIDENCE

The following points highlight and summarise several gaps in the evidence base in relation to inequalities in GI infections in high income countries. The work in this thesis aims to address the gaps identified, to enhance current understanding of inequalities in the consequences of GI infections.

- Firstly, there is some evidence that inequalities in healthcare presentation for GI infections are apparent, yet the mechanisms that might explain these inequalities are poorly understood. All three studies presented in this thesis seek to address this gap in the knowledge base.

- The association between SES and the risk of acquiring a GI infection remains unclear and conflicting findings showing positive and negative social gradients have been observed by several studies that have sampled GI infection cases via population-based surveys. There appears to be no obvious explanation for the contrasting findings observed, however there are many sources of clinical and methodological heterogeneity amongst the studies. A systematic review is therefore warranted to summarise, organise and make sense of the contradictory findings observed in the literature. Study 1 of this thesis seeks to enhance understanding of inequalities in GI infection risk by comparing studies that have identified GI infection cases via population-based surveys, healthcare and laboratory records, using systematic literature review and meta-analytic methods.

- Little is known about the extent of inequalities in IID severity, but there is convincing evidence to suggest that disease severity is a predictor of primary care presentation for IID. A very small number of studies have found inequalities in measures of IID severity such as the duration of illness, and these studies have exclusively focused on paediatric populations. Study 2 of this thesis seeks to address this gap in the evidence base, by exploring inequalities in symptom severity for all age groups, using data obtained from the largest and most up-to-date population-based survey of IID conducted in the UK (the IID2 study).

- Only two previous studies (conducted in the UK and the Netherlands) have investigated the relationship between SES and sickness absence due to IID, and these studies have produced inconsistent results. Neither study controlled for potential confounding variables other than age. Study 2 of this thesis seeks to enhance understanding by

investigating the association between SES and sickness absence using IID2 study data, and by examining whether differences in IID severity can explain any association between SES and sickness absence due to IID.

- There is some evidence to suggest that those of lower SES compared to high are more likely to be hospitalised with a GI infection, although not all studies have observed this association. Inconsistent findings have been observed in the UK, and these studies have tended to aggregate data over large geographical areas, and perform univariate analyses only. Study 3 of this thesis provides an up-to-date assessment of the relationship between SES and hospitalisation for IID in England, and examines the effects of several neighbourhood-level characteristics on the relationship, thereby advancing the literature by taking a more comprehensive approach. Inequalities in both admission rates and the duration of admissions are investigated, using small areas to aggregate data.

- Lastly, the majority of studies that have explored inequalities in healthcare presentation for GI infections, have analysed either paediatric populations only, or individuals of all ages combined. Measuring the association between SES and the risks and consequences of GI infections for adults and children separately, may reveal useful insights that enhance understanding of inequalities in GI infections throughout the life course. Thus, stratification of results by child and adult age groups is featured in the studies of this thesis.

Chapter 3

Methods

In this chapter, I describe the methods for each of the three studies featured in this thesis: a systematic literature review and meta-analysis (Study 1), a cross-sectional analysis of the population-based IID2 study dataset (Study 2), and an ecological cross-sectional analysis of routinely collected HES data (Study 3). The rationale behind the choice of method is discussed.

3.1 STUDY 1

Study 1 was a systematic literature review which aimed to assess the relationship between SES and the risk of symptomatic GI infections in high income countries. Objectives were to evaluate possible sources of heterogeneity in effect estimates reported in the literature. The following section describes the development of the research question, the search strategy and inclusion/exclusion criteria used to identify relevant literature, the quality appraisal tool used to assess bias and reliability of the included studies, and the methods used to analyse and synthesise the results of the studies.

Systematic literature reviews

The purpose of the literature review is to assess and critically summarise the literature on a particular topic (Blaxter, Hughes and Tight, 2010). The methods to conduct a literature review involve searching for and selecting relevant literature, appraising this literature to determine validity and reliability, and analysing and synthesising the reported results (Bettany-Saltikov, 2012; Aveyard, 2010). By comparing and synthesising the results of several studies, reviews can establish the generalisability and consistency of an effect (Greenhalgh, 2010), and novel perspectives and patterns of evidence may be identified (Polgar and Thomas, 2008).

The reliability of the 'traditional' or narrative literature review has come into question due to the ease at which this type of review can be influenced by the researcher's views or preferences (Khan et al., 2011). Narrative reviews in general do not involve comprehensive searches, provide justification for exclusion criteria, or differentiate between good and poor quality studies (Hicks, 2009). The use of such non-standardised, subjective methods can introduce bias into a review and produce misleading conclusions (Greenhalgh, 2010). Systematic reviews differ from traditional reviews and commentaries, in that they are

conducted using explicit, systematic and reproducible methods (Khan et al., 2003). These methods increase the reliability and validity of the conclusions drawn. Additionally, systematic reviews are undertaken to answer focussed research questions, whereas traditional reviews tend to address topic areas (Gough, Oliver and Thomas, 2012).

Research question

The first step in conducting a systematic review is developing a structured, unambiguous research question (Khan et al., 2003). To achieve this, the PICO framework was used which is a method of phrasing questions using four key elements: the population under investigation; the intervention or exposure being considered; the comparison exposure; and the outcome of interest (Richardson et al., 1995). The PICO framework was designed to produce relevant, focused and precise questions (Richardson et al., 1995). It has been suggested that the development of precise research questions enables the researcher to search for evidence with which to answer the question more efficiently (Eldredge, 2000).

The research question for the systematic review was:

- For individuals from high income countries, is lower SES compared to high associated with a higher incidence or prevalence of GI infections?

Population

Individuals, of any age or gender, from high income countries were included. A high income (developed) country was defined as being a member country of the Organisation for Economic Co-operation and Development (OECD). The OECD aims to continually monitor the economic developments of its 34 member countries and provides policy recommendations to help governments tackle poverty through economic growth and stability (OECD, no date).

Exposure

The exposure of interest was lower compared to higher SES, measured at the individual or aggregate level by income, education, occupation, employment or deprivation of area of residence.

Outcome

The outcome of interest was the incidence or prevalence of sporadically occurring symptomatic GI infections measured using population-based surveys, routine surveillance systems, laboratory data, GP presentation data or hospitalisation data, and included syndromic definitions of GI infections without a laboratory diagnosis.

Inclusion and exclusion criteria

Adhering to clearly defined inclusion/exclusion criteria during the screening process can add rigour to a review, and minimises the possibility of selection bias (McDonagh et al., 2013). Therefore, before the searching commenced, several inclusion and exclusion criteria were developed to screen the search results systematically and ensure that only relevant articles were included in the final review. The criteria are displayed in Table 3.1.

Observational studies (cross-sectional, ecological, case-control, cohort [prospective and retrospective]) reporting quantitative results and analysis of empirical data on the prevalence or incidence of any symptomatic GI infection by SES, in a representative population sample were included. Studies that used representative population samples were selected to improve the external validity of the review results. SES could be measured by occupation, income, education, employment or deprivation at the individual or aggregate level. Studies conducted in high income countries (defined as being a member country of the OECD), written in or translated into English, that reported on human subjects and used data collected after 1980 were included. Restricting to publications using data from 1980 onwards ensured that the results were as relevant as possible to the present day. For countries that joined the OECD after 1980, data collection must have occurred after the date the country became a member of the OECD. To glean as much information as possible on the relationship between SES and GI infections, studies which analysed the same cases were included if they analysed different exposures or outcomes. Where more than one study analysed the same cases using the same outcomes and exposures, only one study was included based on the study with the greatest focus and amount of information on the relationship between SES and GI infection risk.

Studies not meeting the above criteria, including case studies, case series, literature reviews, studies conducted solely in a specific population subgroup without a general population comparator group, or studies conducted in institutional settings such as nurseries, hospitals or the military were excluded. Since the outcome of interest was sporadically occurring

symptomatic GI infections, studies that reported on outbreaks of GI infection, or asymptomatic infections only were excluded. Additionally, studies that analysed travel associated illness only were excluded, since the illness may have originated from a non-OECD country. Non-English language studies were excluded due to time limitations and the costs of translating studies.

Table 3.1 Inclusion and exclusion criteria

Inclusion criteria
1. Studies quantitatively measuring the prevalence or incidence of any symptomatic GI infection in a representative population sample
2. Studies quantitatively measuring SES at an individual or aggregate level by occupation, income, education, employment or area deprivation
3. Studies reporting a quantitative association between the first two inclusion criteria, i.e. reporting an association between GI infection risk and SES
4. Studies written or translated into the English language
5. Studies reporting on human subjects
6. Subjects selected from the populations of countries that are members of the OECD, reporting data after 1980 or the date that they became a member of the OECD
7. Studies reporting on data collected after 1980
8. Observational studies
Exclusion criteria
1. Unrepresentative population sample
2. Outbreak reports
3. Studies analysing travel related cases only
4. Review studies
5. Case reports

Search strategy

The results of any review can be biased if relevant literature is missed and subsequently omitted from the review (Lefebvre et al., 2011). A broad search strategy was adopted to ensure, as far as possible, that all relevant literature was identified. Greenhalgh and Peacock (2005) recommend using a variety of methods to enhance a search strategy. Therefore, three stages of searching were performed: 1) database, 2) grey literature, and 3) reference list searching. Employing techniques such as the hand searching of reference lists in conjunction with electronic database searching is important because the indexing of articles in databases is not always accurate or complete (Khan et al., 2011). The three stages are described below.

Stage 1: Database search

The systematic searching of electronic databases formed the initial search, which was performed on the 13th October 2015. It is unlikely that a single search will yield all of the literature that is required (Ross, 2012), and therefore three databases were searched. The decision about which databases to search was made using a list of health sciences databases, provided by the University of Liverpool (University of Liverpool, 2016). The recommended databases were selected; MEDLINE (Ovid), Scopus and Web of Science Core Collection. These were considered most relevant to the research question and likely to yield the highest number of relevant papers since they indexed an extensive number of citations related to medicine and life sciences. Databases which were broad in scope were preferred over those with niche topics, again to yield the highest number of relevant papers. The advice of a university librarian was sought to confirm that the final list of chosen databases was sufficient to retrieve most of the relevant articles, and to check that prominent databases had not been omitted.

Search terms were developed with which to search the databases (Appendix 4). The three main constituents of the research question; 'socioeconomic status', 'gastrointestinal infection' and 'high income countries', were used to develop the search terms. Ultimately, the GI infection terms were selected because they represented the main GI pathogens known to cause the greatest burden to public health in the developed world. Whilst not exhaustive, the list was intended to provide a broad spectrum of bacterial, viral and protozoal infections.

The terms 'socioeconomic status' and 'gastrointestinal infection' were entered into Roget's Thesaurus online (Thesaurus.com, no date), to identify as many synonyms as possible. Additionally, the thesaurus in MEDLINE was used to identify relevant synonyms, by mapping and inspecting the tree for each term using the 'search tools' function. Relevant terms mentioned in articles identified in a preliminary search of the literature were also added. Countries featuring on the list of countries in the OECD were added as individual search terms.

The search terms for MEDLINE were developed initially. Where possible, terms were exploded to broaden the search. Terms were added as keywords if they could not be exploded or if the exploded terms were not relevant to the research question. Truncation and proximity operators were also applied as necessary to broaden the search.

For consistency and to ensure a systematic process, the exact same terms were used for Scopus and Web of Science Core Collection, however as the functionality of each database was different, the terms needed to be adapted for correct use in each. Specifically, the terms contained within the exploded terms in MEDLINE, needed to be added as individual search terms in Scopus and Web of Science Core Collection and phrases needed to be indicated with quotation marks. Additionally, the proximity operators differed for each database. In Scopus and Web of Science Core Collection, each term was searched for within the title, abstract and keywords of the documents contained in each database.

The databases had certain filters which narrowed down the number of results into more manageable amounts for screening. Where available, filters for English language, human subjects, publication year and document type, were applied to the results within each database. These filters were chosen because they directly related to the inclusion criteria. In MEDLINE, there was no available document filter and therefore the results were limited to studies that involved human subjects that were published in English since 1980. In Scopus, the results were limited to studies that had 'human' or 'humans' as keywords that were published in English since 1980, and that were categorised as one of the following documents: article, conference paper, letter, short survey or undefined. In Web of Science Core Collection, there was no human subject filter, and so results were refined to include studies published in English since 1980, that were categorised as articles, proceedings papers, book chapters or letters. Tables were completed showing the number of results for each database search. This was intended to act as evidence of the systematic search methods used in the review, and allows the reader to assess the rigor and validity of the conclusions drawn.

The references remaining after the filters were applied in each database were then exported to EndNote reference managing software. In EndNote, the references from the three databases were combined and duplicates were identified and removed using the 'find duplicates' function and manually by each reviewer. The remaining references were screened for inclusion.

Stage 2: Grey literature search

In addition to the database search, a search of the grey literature was performed. The terms 'gastroenteritis OR "gastrointestinal infection" OR diarrhoea OR diarrhea AND "socio economic" OR socioeconomic OR "social class" OR deprivation' were entered into the Google internet search engine and the Google Scholar search application. Google Scholar

was searched from 1st January 1980 to 31st December 2015, with citations included and patents excluded. Google was searched over the same time period, with Google Instant results turned off. The first 100 results in order of relevance were screened for inclusion.

Stage 3: Reference list search

The third stage involved hand searching the reference lists of studies selected for inclusion in the review to identify potentially relevant articles that were not captured by electronic searching.

Screening the search results

The results identified by the three search stages were screened for inclusion using the inclusion and exclusion criteria (Table 3.1). Titles and abstracts were screened independently by two reviewers (Natalie Adams and myself) to ensure consistency in the application of the inclusion and exclusion criteria. Any discrepancies were discussed and re-examined until we reached an agreement. The full text for studies deemed relevant after title and abstract screening were retrieved and reviewed in the same way. At each stage, EndNote software was used to record and organise the studies.

Access to the sample was attained via the University of Liverpool's and PHE's collection of electronic journals. Studies were also accessed through journals freely available via the internet. Where full texts were not available, they were sought via institutional library sharing agreements.

Quality appraisal of studies

Assessing included studies for the risk of bias is an essential component of a review, since certain biases can have substantial effects on the results of a study (Higgins et al., 2011). Critical appraisal tools can be used to provide a systematic framework for the quality appraisal, to ensure as far as possible that all studies are reviewed with equal rigour (Aveyard, 2010). Since different study designs can be prone to certain biases, strengths and weaknesses (Burls, 2009), it was decided that a study design specific appraisal tool should be used to achieve a more rigorous and accurate assessment of the biases which may have affected the results of each study.

The Liverpool University Quality Assessment Tool (LQAT) was used for this review, which allowed the methodological quality of the studies to be assessed using a tool specific to each study design (case-control, cohort and cross-sectional) (Pope, 2015). The LQAT has been used in previous systematic reviews (Rehfuess et al., 2014; Puzzolo et al., 2013) and has been independently evaluated against other quality assessment tools (Voss and Rehfuess, 2013). It incorporates a star rating system to assess and qualify absence of bias, misclassification and confounding. A benefit of using this tool was that the results of the quality assessment could be quantified into a score for each study based on the star rating system. For each study design, the quality scores were converted into tertiles, whereby three approximately equally sized groups were created according to the distribution of the raw quality scores, to denote high, medium and low quality studies. The ability to categorize the quality of the studies in this manner permitted sensitivity analyses to be performed excluding low quality studies.

An additional benefit of using the LQAT was that minor adaptations could be made to the categories of the exposure and outcome assessment questions to better suit the research question (Appendix 4). The methods used to measure the exposure were assessed according to whether SES was measured and analysed per group (i.e. not determined individually); measured per individual but aggregated to an area-based measure; or measured and analysed per individual. The methods used to measure the outcome were assessed according to whether GI infections were measured via self-reported symptoms ascertained retrospectively; self-reported symptoms ascertained prospectively; health record/physician diagnosis based on symptoms; or laboratory confirmation.

To ensure a systematic process and to add rigor to the review, the quality assessment of the studies was conducted independently by two reviewers (Natalie Adams and myself), and any discrepancies between assessments were discussed and re-examined.

Data analysis and synthesis

Hart (1998) describes data analysis as the systematic dissection of research articles, enabling comparisons and contrasts to be drawn. Once these comparisons and contrasts have been identified, they should be questioned to explain why such relationships or differing viewpoints exist (Ross, 2012). To organise the data and facilitate comparison, tables were created by extracting data from each study into a standardised Microsoft Excel spreadsheet. Extracted data included: aim/hypothesis; study design; level of analysis; country; sample

size; age; age category; type of GI infection; GI infection method of measurement and data source; measure of SES; SES method of measurement and data source; confounding variables controlled for; statistically significant results; non-significant results; conclusions; and quality assessment. Extracted data were checked for accuracy by at least one other reviewer. For studies where quantitative data were reported in text form only, authors were contacted to obtain the relevant data, however no further information was obtained.

Once data were extracted, subgroup analyses were performed on study design factors and potential modifying factors of the association between SES and GI infection risk that were identified a priori based on knowledge of the subject matter, including: pathogen type (based on mode of transmission); country (based on climate and level of development); age of participants; measure of SES; and the methods used to sample GI infection cases. It was considered that these factors may have been associated with SES or GI infection risk. Identifying these subgroups a priori minimised the potential of drawing false positive conclusions due to 'data dredging' or performing multiple analyses of the data (Baker et al., 2009; Thompson and Higgins, 2002). The scientific rationale for the choice of potential modifying factors is presented in Table 3.2. It was considered that information on these factors would be available for most studies, which would help maximise the power to detect modifying effects in the meta-regression (as described below) (Hempel et al., 2013).

Studies that analysed specific pathogens were assigned one of four categories based on the predominant mode of transmission of the pathogen, i.e. foodborne, waterborne, environmental or person-to-person. Studies were assigned one of three age categories (children <18 years; adults ≥18 years; mixed ages) based on the age of the participants under investigation, and one of four categories indicating the methods used to sample cases, i.e. population-based surveys, laboratory records, hospital admissions or GP presentations. In addition, studies were categorised into those which used area-level (e.g. IMD) or individual-level measures of SES.

The countries within which the studies were conducted were ranked by relative level of development using the Human Development Index (United Nations Development Programme, no date), and climate zones were assigned based on the Köppen system (Met Office, 2015). Each country's predominant climate zone was chosen. For larger countries with multiple climate zones (e.g. Australia, USA) the predominant climate zone relating to the study sample was chosen. Three climate categories were investigated: Temperate/Mediterranean, Arid and Snow.

Table 3.2 Potential modifiers of the association between SES & GI infection risk

Age	Consistent evidence shows that age is an important risk factor for GI infections (LaRocque and Calderwood, 2015).
Pathogen type	Certain pathogens such as norovirus occur far more frequently in the community, compared to pathogens such as STEC (Tam et al., 2012c), indicating that risk of infection can vary by pathogen.
Country: climate	Several studies have found that weather and climate factors such as temperature, humidity and rainfall, can influence the incidence of GI infections (Onozuka, 2014; WHO, 2004).
Country: development	A country's level of development has been found to be a risk factor for GI infection risk; countries with lower levels of development compared to higher, tend to have higher incidences of infection (Fletcher, McLaws and Ellis, 2013).
SES measure	Individual measures of SES, such as occupation, education and income, each capture a distinct aspect of SES but may correlate with each other and other measures of SES (Shavers, 2007). Area-level SES measures may capture contextual factors shared by communities, however studies have found that these measures sometimes do not correlate well with individual measures (Pardo-Crespo et al., 2013; Shavers, 2007). Since different measures capture different aspects of SES, the association between SES and GI infection risk may be modified by the type of SES measure used.
Source of cases	The incidence of GI infections measured using laboratory records, GP presentations, hospital admissions and population-based surveys is around 0.2%, 2%, 0.6% and 27%, respectively (Tam et al., 2012b; Olowokure et al., 1999). This indicates that the risk of GI infection can vary according to sources from which cases are sampled.

Both harvest plots and meta-analyses were used to synthesise the diverse body of evidence. Not all of the data could be combined in meta-analysis, and therefore harvest plots were created to compliment the meta-analysis, allowing the results of all of the studies to be captured.

Harvest plots

Harvest plots were created for each subgroup; displaying and summarising the results of the studies and the subgrouping graphically (Ogilvie et al., 2008). Harvest plots are matrices that are used as visual aids to summarise the findings of clinically and methodologically heterogeneous studies (Ogilvie et al., 2008). Harvest plots were originally devised to illustrate visually differential impact by SES and to show where the weight of evidence lies in terms of inequalities. Additionally, the matrix format allows the results of studies to be partitioned into subgroups of interest, and comparisons between the subgroups can be assessed. An inclusive strategy was used for the harvest plots, allowing all studies to be

captured. All point estimates between SES and GI infections were included in the harvest plots, including estimates that were mentioned in text form only. Additionally, more than one estimate from the same study could be included, for example, where studies provided multiple estimates using different SES measures. This inclusive strategy was in contrast with the strict inclusion criteria required for the meta-analysis (as detailed below). By including all of the studies, harvest plots provided a more complete overview of the results compared to the meta-analysis, and using both narrative and statistical synthesis methods increased the reliability of the results. In this respect the harvest plots made a valuable addition to the presentation and interpretation of results by SES.

The harvest plots partitioned the point estimates between SES and risk of GI infection into three categories: estimates that indicated GI infection risk was lower amongst those of lower SES; estimates that indicated GI infection risk was higher amongst those of lower SES; and estimates that indicated there was no relationship between GI infection and SES. Each point estimate was represented by a single bar and the height of the bars were used to indicate the quality score of the studies from which the estimates originated (low, medium or high quality), so that the strength of the evidence could be determined and greater weight given to conclusions drawn from the most methodologically robust and reliable studies. The findings from the harvest plots were used to inform the methods used in the meta-analysis, and lead to potential explanations for the contrasting findings observed in the literature.

Meta-analysis and meta-regression

Meta-analysis was also used to synthesise the findings of the studies. Meta-analysis is a statistical technique which synthesises the results from several studies to form a more statistically powerful result (Fink, 2010). Meta-analysis can therefore be used to summarise the results of many studies by providing a precise, overall estimate of an effect. However, meta-analysis of observational studies has certain limitations which should be considered. Compared to randomised-controlled trails, the results of observational studies are prone to the effects of bias and confounding, and thus statistically combining the results of observational studies can produce a statistically precise but nonetheless spurious estimate of an effect (Egger, Smith and Schneider, 2001). Therefore, it has been suggested that the purpose of meta-analysis of observational studies is to explore potential reasons for heterogeneous risk estimates across studies, rather than to obtain an overall summary statistic for the combined studies (Kheifets et al., 1995). Random-effects meta-regression (Thompson and Sharp, 1999; Berkey et al., 1995) and subgroup meta-analyses were therefore performed to investigate potential modifying factors (identified a priori) of the relationship between

SES and GI infection risk, and this process was guided by the findings from the harvest plots. Statistical heterogeneity was assessed by applying the I^2 statistic with values of 30 to 60%, 50 to 90% and 75 to 100% used to denote moderate, substantial and considerable levels of heterogeneity, respectively (Deeks et al., 2011).

Meta-analyses were conducted using R statistical software (version 3.3.1) using an inverse variance random-effects model on combined results. Using this method, a pooled effect estimate is calculated as a weighted average of the estimates provided in the individual studies, where the weight given to each study is the inverse of the variance of the study's effect estimate (Deeks et al., 2011). In this way, more weight is assigned to studies that yield more precise estimates of the effect (Borenstein et al., 2010). The random-effects model assumes that the studies are estimating different, yet related effects, and thus incorporates a measure of the extent of variation among the effect estimates observed between the different studies (Deeks et al., 2011). In contrast, the fixed-effect model assumes there is one true effect that underlies all of the studies; therefore fixed-effect models do not take into account between-study variance, and consequently the variance of the pooled effect estimate is usually smaller under the fixed-effect compared to the random-effects model (Borenstein et al., 2010). Since this review sought to capture a broad range of studies (studies conducted in different populations, using different measures of the exposure and outcome of interest), using a random-effects model was deemed an appropriate choice.

Random-effects meta-regression was used to investigate whether various clinical or methodological differences among the studies could explain any of the statistical heterogeneity in the study estimates. In a meta-regression, a dependent variable (the effect estimate) is predicted by one or more independent variables (the potential effect modifiers that might influence the size of the effect estimate) (Deeks et al., 2011). The studies are the units of analysis, and they are weighted so that more precise studies have more influence in the analysis (Baker et al., 2009; Thompson and Higgins, 2002). Similar to the reasoning outlined above, a random-effects model was considered more appropriate than a fixed-effect model, since the former adjusts for between-study variability and it would be unreasonable to assume that there would be no residual between-study variability even after accounting for multiple sources of heterogeneity in a meta-regression analysis (Baker et al., 2009).

In contrast to the inclusive strategy used for the harvest plots, it was necessary to apply stricter criteria for the inclusion of studies in the meta-analysis. Where included studies analysed the same cases or provided more than one estimate for the relationship between SES and GI infection (for example providing multiple estimates based on different measures

of SES), only one estimate was retained in the meta-analysis to avoid the double counting of cases. Where studies provided estimates for more than one SES measure, the most commonly used SES measure across all of the studies (education level) was chosen for inclusion in the meta-analysis. Likewise maternal education was chosen over paternal education as a greater number of studies reported maternal education. Where one estimate was based on a smaller subset of cases as another reported estimate from the same study, the estimate based on the larger number of cases was included only. Estimates that were statistically adjusted for the effects of potential confounding variables, such as age, were chosen over univariate estimates where both types of estimates were provided within a study. This was to reduce the potential effects of confounding since all included studies were observational.

Eleven studies provided more than one estimate but the cases used for each estimate were considered to be independent of each other so all estimates were included in the meta-analysis. Studies that belonged to this category were those that provided estimates for children and adults, or estimates for two separate geographical areas, as well as studies that provided estimates for different pathogens since it was assumed that it would be unlikely for a case to be infected with more than one pathogen. However, a potential issue when including multiple estimates from a single study in a random-effects meta-analysis is that the within-study variability of different estimates would be treated as between-study variability, and therefore studies with multiple estimates would have received a disproportionately high weight in the pooled estimate. Therefore fixed-effect meta-analyses were used to combine point estimates from the same study, allowing these pooled estimates to be combined with the remaining studies using random-effects meta-analysis (Deeks et al., 2011).

Meta-analysis of dichotomous outcomes was performed. Odds ratios, relative risks, hazard ratios and rate ratios were combined. Odds ratios when interpreted as relative risks, always overstate any effect size, however they are not greatly dissimilar to relative risks when the risk of the outcome under investigation (in this case the disease incidence) is approximately <20% (Davies, Crombie and Tavakoli, 1998). Based on UK data, the incidence of IID measured using laboratory records, GP presentations, hospital admissions and population-based surveys is around 0.2%, 2%, 0.6% and 27%, respectively (Tam et al., 2012b; Olowokure et al., 1999). Overall, it was therefore considered appropriate to combine odds ratios and relative risks, however sensitivity analyses were performed to check the appropriateness of making this assumption for odds ratios derived from population-based surveys, as described below.

Where the SES exposure was a categorical variable, the high SES category was taken as the baseline, and thus the direction of the association was GI infection risk for low compared to high SES. Where studies presented the relationship between high SES compared to low, the inverse/reciprocal of the results were taken in order to present the results as low compared to high SES.

Studies that did not directly report a point estimate for the association between SES and GI infection risk, were investigated to determine whether a point estimate could be calculated from the raw data reported in the study. For studies that reported the prevalence of GI infections in low and high SES groups, an unadjusted odds ratio and standard error were calculated using the following formula:

	Number with GI infection	**Number without GI infection**
Low SES	*a*	*b*
High SES	*c*	*d*

Odds ratio = $\frac{ad}{bc}$

Standard error (of log odds ratio) = $\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}$

(LaMorte, 2017)

For studies that reported the cumulative incidence of GI infections in low and high SES groups, an unadjusted relative risk with standard error were calculated using the formula:

	Number with GI infection	**Total number subjects**
Low SES	*a*	N_1
High SES	*c*	N_0

Relative risk = $\frac{a/N_1}{c/N_0}$

Standard error (of log relative risk) = $\sqrt{\frac{1}{a}-\frac{1}{N_1}+\frac{1}{c}-\frac{1}{N_0}}$

(LaMorte, 2016a; LaMorte, 2016b)

Finally, for studies that reported the incidence rate of GI infections in low and high SES groups, an unadjusted rate ratio with standard error were calculated using the formula:

	Number with GI infection	Person-time at risk
Low SES	a	Y_1
High SES	c	Y_0

Rate ratio = $\frac{a/Y_1}{c/Y_0}$

Standard error (of log rate ratio) = $\sqrt{\frac{1}{a}+\frac{1}{c}}$

(LaMorte, 2016a; LaMorte, 2016b)

For studies that reported a point estimate and a 95% confidence interval only, the standard errors were calculated from the 95% confidence intervals, using the formula:

Standard error = (log(upper confidence interval) - log(lower confidence interval)) / 3.92

(Higgins and Deeks, 2011)

Some studies analysed SES as a continuous variable, presenting risk ratios per unit increase in SES exposure, rather than risk ratios for low compared to high SES categories. In order to combine these two types of results, information on the distribution of the SES exposure reported in the study, subject-specific knowledge and reasonable assumptions (where required) were used in order to estimate a plausible low versus high SES comparison for studies that analysed SES as a continuous variable. For a log risk ratio of x per unit increase in SES exposure, and a low compared to high SES difference of d, the risk ratio for low compared to high SES was obtained using the formula below, with a similar formula for the lower and upper confidence intervals:

Risk ratio = $\exp(xd)$

For instance, Weisent et al. (2012) conducted an ecological study and used a continuous SES measure; the proportion of the population with no high school diploma. This continuous SES variable had a reported standard deviation (SD) of around 10%. We therefore based our low

versus high SES comparison on a 20% difference in proportion of the population with no high school diploma (or +1 SD versus -1 SD below the mean). Similarly, for studies that analysed SES as a continuous variable on a scale of 1–10 (e.g. Spencer et al., 2012), we used the comparison of 2 versus 9, or a difference of 7, corresponding to a comparison between the lower end of the SES scale compared to the higher. This approach may of course lead to an exaggeration or understatement of the risk ratio for low compared to high SES, depending on whether studies that report low compared to high SES tend to compare a narrower or wider range of exposure than that assumed for the value of *d*. However, this is a general problem with combining estimates of low compared to high SES exposure, as the reported risk ratios may be more or less extreme for some studies, depending on the categorisations employed. For example, one study may compare the bottom decile of an exposure versus the top decile, and another the bottom third versus the top third of the SES exposure distribution.

Finally, three studies (Bemis, Marcus and Hadler, 2014; Beale et al., 2010 and Fein et al., 1995), did not report point estimates for the association between SES and GI infection risk, and provided only the total sample size and the proportion of GI infection cases in each SES category (or the equivalent information). To calculate point estimates for these studies, it was assumed that the total number of individuals in each SES category were identical.

Sensitivity analyses

A number of sensitivity analyses were performed to check the robustness of the results. In particular, it was important to check whether the assumptions that were made when calculating point estimates, and when combining odds ratios and relative risks (particularly for population-based surveys), had influenced the main results of the review. Subgroup meta-analyses were repeated excluding: studies where we calculated point estimates from the raw data; population-based surveys reporting odds ratios; studies classified as being of low quality; studies that did not statistically control for potential confounding variables; and studies that reported point estimates that were adjusted for other SES measures either statistically or by matching in case-control studies.

Some of the studies that provided point estimates for different pathogens, used the same individuals within the denominators when calculating risk ratios for each pathogen, and therefore non-cases may have been counted more than once in the meta-analysis. Deeks et al. (2011) recommend combining groups from multi-estimate studies to create a single pair-wise comparison, however not all studies provided the data required. Therefore, to assess the

impact of double counting non-cases, sensitivity analysis was performed whereby only one estimate from each multi-pathogen study was entered into the meta-analysis. The estimate for the pathogen with the largest number of cases was chosen within each multi-pathogen study.

Finally, publication bias and small study effects were assessed using a funnel plot. Reporting and publication biases can arise if non-significant findings are less likely to be reported in studies, or if studies reporting non-significant findings are less likely to be published and included in a review (Sterne et al., 2011a; Peters et al., 2008). Funnel plots can be used to assess whether these biases may have affected the results of a meta-analysis. Effect estimates of the included studies are plotted against a measure of each study's size or precision, such as the standard error (Sterne et al., 2011a). Effect estimates from smaller studies tend to be more variable than larger studies, and therefore estimates from smaller studies scatter more widely at the bottom of the plot creating a 'funnel' shape (Sterne et al., 2011b; Peters et al., 2008). Asymmetry between points within the plot may indicate the presence of publication bias. However, asymmetry can also be caused by factors that are associated with both the size of a study and the effect estimates that are produced, for example if lower quality studies tend to be smaller in size (Peters et al., 2008). Contour-enhanced funnel plots can help to distinguish between asymmetry caused by publication bias and asymmetry caused by other factors, by displaying contours/regions that define the statistical significance of the study estimates (Peters et al., 2008). If asymmetry occurs only in the region of statistical non-significance, this supports the possibility that publication bias may be present, however if asymmetry occurs in regions of statistical significance, factors other than publication bias may be at play. For this review, contour-enhanced funnel plots were created using the R metafor package (The metafor Package: a meta-analysis package for R, 2016).

Ethical considerations

This was a secondary piece of research and therefore did not require formal ethical approval. Despite this, there was an ethical obligation to conduct the review as objectively as possible, taking measures to minimise bias and critique research fairly (Hart, 1998).

3.2 STUDY 2

Study 2 was a secondary analysis of data collected in the IID2 study, which aimed to explore the association between SES and measures of self-reported IID symptom severity and sickness absence due to IID. The following section describes the IID2 study as the data source, the study design used to address the aims of the research, the selection of dependent and independent variables, and the methods used to analyse the data.

Data source

The IID2 study was a population-based survey of IID, conducted across the UK in 2008–9. It was commissioned by the FSA and Department of Health to determine if the incidence of IID had changed since the mid-1990s, and to re-calibrate national surveillance data (Tam et al., 2012b). The IID2 study contained a number of components including a prospective population-based Cohort study and a GP Presentation study, involving 88 and 37 practices across the UK, respectively (Tam et al., 2012b).

For the Cohort study, a study nurse in each practice created a randomised list of 800 registered patients which was reviewed by a general practitioner who excluded patients meeting the exclusion criteria (Table 3.3) (Tam et al., 2012b). The remaining participants were sent a postal invitation letter and information about the study (O'Brien et al., 2010). If interested, patients were given an appointment with a study nurse who would explain the study in detail and obtain written informed consent if appropriate. All members of the cohort completed a baseline questionnaire containing questions on socio-demographic factors such as age, sex, ethnicity and occupation.

Participants were followed up weekly for one year, to determine whether they had experienced an episode of IID. Cases self-reported IID based on their own definition/perception of an episode of IID. Incident cases completed a symptom questionnaire containing questions on the symptoms experienced, symptom duration, absenteeism and other relevant items, and provided a stool sample.

Table 3.3 Exclusion criteria for IID2 Cohort and GP Presentation studies

1	Patients with terminal illness
2	Patients whose first language was not English and for whom a suitable interpreter was not available
3	Patients with severe mental incapacity
4	Patients with non-infectious causes of diarrhoea or vomiting: Crohn's disease, ulcerative colitis, cystic fibrosis, coeliac disease, surgical obstruction, excess alcohol, morning sickness and, in infants, regurgitation

Source: O'Brien et al. (2010)

The GP Presentation study aimed to recruit all cases of IID who met the case definition but not the exclusion criteria (Table 3.3), and consulted a healthcare practitioner in person or by telephone, or were seen by an out-of-hours service provider for their illness (Tam et al., 2012b). Contact with NHS Direct was not included.

Healthcare practitioners provided eligible patients with an information sheet about the study during the consultation. Study nurses invited interested patients to attend a baseline interview where the study would be explained in detail and written informed consent obtained if appropriate (Tam et al., 2012b). Each participant completed a baseline questionnaire containing questions on socio-demographic factors, and also questions relating to the illness itself.

The case definition used in the Cohort study and the GP Presentation study differed in that cases self-reported IID based on their own definition/perception of an episode of IID for the Cohort study, whereas healthcare practitioners applied the case definition, as per Table 3.4, for the GP Presentation study. The case definition used in the GP Presentation study was therefore perhaps more specific than the case definition used in the Cohort study. The exclusion criteria (Table 3.3) were applied in both studies by healthcare professionals.

Table 3.4 Case definition for IID2 GP Presentation study

Case definition of IID
People with loose stools or clinically significant vomiting lasting less than two weeks, in the absence of a known non-infectious cause, preceded by a symptom-free period of three weeks. Vomiting was considered clinically significant if it occurred more than once in a 24-hour period and if it incapacitated the case or was accompanied by other symptoms such as cramps or fever.

Source: O'Brien et al. (2010)

Population

The sampling frame for the IID2 study, was 912 general practices in the UK that were part of the Medical Research Council General Practice Research Framework, and general practices from Primary Care Research Networks in England, Wales, Scotland and Northern Ireland (O'Brien et al., 2010). As mentioned, 88 practices took part in the Cohort study, and 37 of these practices also took part in the GP Presentation study (Tam et al., 2012b). A very small number of participants were recruited to both studies (O'Brien, 2015), and it was not possible for me to select and exclude these participants from the available datasets.

Overall, 7033 participants were recruited to the Cohort study, however two participants withdrew consent during the study, consent was not verified for 11 participants, and 184 participants, who were recruited close to the end of the study, did not contribute any follow-up time (Tam et al., 2012b). Therefore, 6836 participants formed the cohort.

Compared with the UK population, Cohort study participants were generally older, with a particular deficit among males between the ages of 15–54 years (Tam et al., 2012b). Individuals of White ethnicity were over-represented, while other ethnic groups were slightly under-represented in the cohort. Those in managerial/professional occupations, as measured by the NS-SEC, were over-represented in the cohort, and participants in intermediate, semi-routine and routine occupations were under-represented. Individuals living in rural areas were over-represented in the cohort compared with the UK census (Tam et al., 2012b).

In total, 2233 patients were referred to the GP Presentation study, and 2203 (99%) were invited to take part. Among those invited to participate, 1254 (57%) attended a baseline interview and were recruited (Tam et al., 2012b). Participation did not vary greatly between males and females. Of those invited, 57.1% of males consented to participate compared to 56.8% of females, however overall a greater number of females were recruited (n=665). Participation was lowest among the 15–24 year age group and highest among the 55–64 year age group for both sexes (Tam et al., 2012b).

To assess under-ascertainment, Read codes for IID were used to search practice databases and identify all IID-related presentations occurring during the same time period as the GP Presentation study (Tam et al., 2012b). With travel-related cases removed, there were six additional cases identified in the practice database for every participant enrolled in the GP Presentation study (Tam et al., 2012b). Under-ascertainment was higher among females than males, and among individuals aged <25 years compared with other age groups. Data on SES

was not collected and therefore assessing under-ascertainment across socioeconomic groups was not possible.

Study design

Secondary data analysis was performed on individuals with IID identified from the GP Presentation study and Cohort study components of the IID2 study. Only IID cases were included in the analysis because data on the outcomes of interest were only collected for cases. A cross-sectional analysis was performed with these IID cases to assess the relationship between SES and measures of IID symptom severity and sickness absence. Cases aged five years or older were included to limit potential misclassification of the more subjective symptoms, such as headache and nausea, in young children (see below details of symptom severity score). Cases of school or working age were used to investigate the sickness absence outcome to improve the interpretation of the results.

Outcomes

Outcome 1: Symptom severity score

The first outcome (dependent variable) of interest was the severity of IID as perceived by each case. Cases were asked to complete a symptom questionnaire to ascertain information relating to the severity of their illness, including the presence/absence of certain symptoms and the duration of the symptoms (Table 3.5). I combined these variables to create a single severity score for each case.

In a previous publication, Tam, Rodrigues and O'Brien (2003) created a severity scoring system for the IID1 dataset by multiplying the presence, duration and severity score for each symptom and adding these product scores together to create an overall score for each patient. This scoring system was used for this analysis, since most of the variables used to derive the score were available within the IID2 dataset due to similarities between the IID1 and IID2 study questionnaires.

Cases could respond 'Yes', 'No' or 'Don't know' when asked about the presence of each symptom. For this analysis, if they responded 'Yes' to the presence of a symptom, they were given a score corresponding to the relative severity of the particular symptom (Table 3.5). If

the case did not experience the symptom, they were given a score of zero for that symptom. The 'Don't know' answers to the symptom present questions were classified as missing.

Cases were also asked about the duration of each of the symptoms: diarrhoea, diarrhoea with blood, vomiting and nausea. The durations were also coded using the scoring system in Table 3.5. The duration and presence scores were multiplied, and then added together with the remaining scores for each symptom, to create an overall single severity score for each case. Using this method, if a case had experienced either abdominal cramps, loss of appetite, fever, cough/runny nose/sore throat, or headache, the duration was assumed to be 1–2 days.

The symptom severity variable was positively skewed and was converted into tertiles, whereby three approximately equally sized groups were created according to the distribution of the severity score; a method which was also employed by Tam, Rodrigues and O'Brien (2003). The boundaries of the tertiles were: mild (severity score 2–9), moderate (severity score 10–15) and severe (severity score 16–40). To reduce the potential for misclassification in the symptom variables, especially nausea, abdominal cramps and headache which may be considered to be particularly subjective, I restricted the analysis sample to cases aged five years and over, and calculated the symptom severity score only for those cases.

Table 3.5 Symptom severity scoring system

	Present		Duration in days			
Symptoms	Yes	No	1–2	3–4	5–6	7+
Diarrhoea	2	0	1	2	3	4
Vomiting	2	0	1	2	3	4
Diarrhoea with blood	3	0	1	2	3	4
Nausea	2	0	1	2	3	4
Abdominal cramps†	3	0	-	-	-	-
Loss of appetite†	2	0	-	-	-	-
Fever (high temperature) †	3	0	-	-	-	-
Cough or runny/blocked nose or sore throat†	1	0	-	-	-	-
Headache†	2	0	-	-	-	-

† Duration data unavailable; a duration score of 1 was applied when calculating the severity score

Outcome 2: Absence from work, school or daily activities

The second outcome of interest was absenteeism due to IID. Cases were asked whether the episode of IID stopped them from going to work, going to school, or carrying out their normal daily activities, to which they could respond 'Yes', 'No' or 'Don't know'. The 'Don't know' responses were re-coded as missing. To improve the interpretation of the

results, the sickness absence outcome was investigated using only cases likely to be of school or working age. Cases of school or working age were defined as those aged five years or older, and up to 60 years for women and 65 years for men, based on the state pension age (ONS, 2005). This was deemed important because absenteeism amongst the population of school/working age may have had a different meaning compared to absenteeism amongst young children and retirees.

Exposures

Primary exposure: National Statistics Socioeconomic Classification

The main exposure (independent variable) of interest was SES which was measured at an individual level using the NS-SEC (ONS, no date). Area level IMD quintiles were also available per participant based on their area of residence. However, since recruitment was conducted via general practices, it was considered that IMD quintiles may have masked SES differences between individuals recruited from the same practice. Thus NS-SEC was chosen as the primary exposure of interest for this analysis.

The NS-SEC is occupationally based and is designed to measure employment relations and conditions of occupations (Rose, Pevalin and O'Reilly, 2005; ONS, no date). Within the analytic version of the NS-SEC, there are eight classes which distinguish different positions based on social relationships within the workplace (ONS, no date). These range from relationships whereby the employee is compensated for their work immediately (e.g. salary) and prospectively (e.g. job security, career advancement), to relationships whereby the employee is provided a wage calculated on the amount of work done or time worked (ONS, no date). The classes also differentiate between large employers and small employers, and the self-employed with no employees (ONS, no date). Classifying those who have never worked and the long-term unemployed in a separate category is optional. The eight classes of the analytic version can be collapsed into five- or three-class nested versions. Individuals classed as full-time students, for whom occupation is not stated or inadequately described or those who are not classifiable for other reasons, are added as 'Not classified' which is a residual category that is excluded when the classification is collapsed into classes (ONS, no date).

For the IID2 studies, the five-class self-coded method was used to derive NS-SEC, which is less accurate but simpler and less expensive than the interviewer-coded method (Tam et al.,

2012b; ONS, no date). Those who had never worked and the long-term unemployed were not separated into a unique category using this method.

To derive NS-SEC, participants were asked four questions about the occupation and employment status of the main-earner in their household. Participants were asked to classify the main-earner's occupation (from eight categories), whether they were an employee or self-employed, the number of employees in their workplace and whether they were a supervisor. Participants were asked to report on the main-earner's current or last main job. In situations where the main-earner was unemployed, retired, looking after home/family or currently sick/disabled, an NS-SEC class was assigned providing that information about their last main job was available. Individuals were assigned the category 'Not classifiable' if information was missing and as such an NS-SEC class could not be calculated. Missing data can be imputed or cases with missing data can be treated as missing (ONS, no date).

Table 3.6 Five-class version to three-class version of NS-SEC

Five classes		Three classes	
1	Higher managerial, administrative and professional occupations	**1**	Higher managerial, administrative and professional occupations
2	Intermediate occupations	**2**	Intermediate occupations
3	Small employers and own account workers		
4	Lower supervisory and technical occupations	**3**	Routine and manual occupations
5	Semi-routine and routine occupations		

Source: ONS (no date)

The five-class NS-SEC version cannot be classed as an ordinal scale, and therefore I re-coded the five-classes to form the nested three-class NS-SEC version which can be assumed to have a hierarchy (Table 3.6) (ONS, no date). The classes from high to low SES represented managerial/professional, intermediate and routine/manual occupations.

Covariates: Age, sex, ethnicity, urban/rural residency, foreign travel

Data were collected on personal characteristics such as the age, sex and ethnicity of the participants. These variables were considered to be potential confounders of the relationship between SES and measures of IID severity, as they may have been related to both SES and disease severity, but were unlikely to be in the hypothetical causal pathway between them

since SES does not determine an individual's age, sex or ethnicity (Figure 3.1). The initial age of each case at the start of follow-up was used. Participants were asked to define their ethnicity from 15 nominal groups; White (UK), White (other), Black (Caribbean), Black (African), Black (other), Indian, Pakistani, Bangladeshi, Chinese, Other Asian, Mixed (White & Black Caribbean), Mixed (White & Black African), Mixed (White & Asian), Other Mixed, Other. Since the vast majority of the participants were of White ethnicity, I re-coded the 15 category ethnicity variable into a binary variable with categories: White and Non-White ethnicity.

Data were also collected on foreign travel in the ten days before disease onset ('Yes' or 'No') and urban/rural residency. The urban/rural residency variable was created by the owners of the data using participants' postcodes which were linked to a defined geographic boundary; a Super Output Area (SOA). The ONS Postcode Directory was used to classify participants' SOA of residence as urban, town or rural (Tam et al., 2012b). For this analysis the urban and town categories were combined to ease the interpretation of the results.

It was considered that these variables could have been related to SES and IID severity, and also may have featured in the hypothetical causal pathway between them (Figure 3.1) (MacKinnon, 2011). For example, SES may determine urban/rural residency as levels of deprivation are generally lower in rural compared to urban areas (Gartner et al., 2008), and there exists a social gradient in holidays/travel abroad (Seaton, 1992). Rural residency may put individuals at increased exposure to certain pathogens, for example STEC which can result in severe sequalae (Byrne et al., 2015), and foreign travel may increase exposure to species such as *Shigella dysenteriae* which is associated with bloody diarrhoea and can induce severe illness (PHLS Advisory Committee on Gastrointestinal Infections, 2004).

Statistical analysis

Cases from the GP Presentation study and Cohort study were combined for this analysis to increase the size of the IID case cohort and enhance the power of the statistical analyses (Biau, Kernéis and Porcher, 2008). The appropriateness of combining cases from the IID2 Cohort and GP Presentation studies was examined by investigating whether the effect of NS-SEC on the outcomes of interest, statistically significantly differed between the Cohort and GP Presentation studies (Appendix 5). All recurrent episodes of IID were removed regardless of the timeframe between episodes. If a case experienced more than one episode

of IID during follow-up, only information related to the first episode was retained to create a sample of independent observations.

Ordinal logistic regression (otherwise known as a proportional odds model or cumulative link model) was used for the symptom severity outcome (Ripley, 2017; McCullagh, 1980). This analysis method was chosen because the symptom severity outcome was an ordinal categorical variable and it was considered beneficial to retain information related to the ordering, as compared to multinomial logistic regression where it is assumed there is no order to the categories of the outcome (UCLA: Statistical Consulting Group, no date-a). For the sickness absence outcome, logistic regression was deemed appropriate due to the binomial distribution of the outcome. Some have cautioned against using logistic regression for modeling cross-sectional data, because this method estimates the odds ratio as an effect measure, which is often misinterpreted as the relative risk when outcomes are common (Lee, Tan and Chia, 2009). Odds ratios do not approximate the relative risk when outcomes are common, instead they likely overestimate the relative risk when >1 and underestimate the relative risk when <1 (Davies, Crombie and Tavakoli, 1998; Zhang and Yu, 1998). Alternative methods for modeling dichotomous outcomes such as log-binomial, Poisson or Cox models with robust variance estimates, may provide better estimates of the relative risk, however these methods can produce biased estimates or estimates with inflated variance resulting in wider confidence intervals, especially when the outcome is common (Reichenheim and Coutinho, 2010; Lee, Tan and Chia, 2009). Therefore, it was decided to use logistic regression since this is the best method for modeling binomial outcomes (Lee, Tan and Chia, 2009), and to make explicit that the odds ratios do not approximate the relative risk when the outcome is a common event.

Ordinal logistic and logistic regression models can be described as generalized linear models (GLMs). GLMs permit the modeling of dependent variables with non-normal distributions e.g. binary and count data (Fox and Weisberg, 2011). The mean of the dependent variable is predicted by a linear function of the independent variables called the linear predictor (Fox and Weisberg, 2011). In GLMs, a link function is used to translate the scale of the mean of the dependent variable to the scale of the linear predictor, i.e. it specifies how the expected value of the dependent variable relates to the linear predictor (Fox and Weisberg, 2011). Both ordinal logistic and logistic regression models typically use a logit link function (see below) which specifies that the conditional mean of the dependent variable is to be constrained to 0 and 1: a binary response (Fox and Weisberg, 2011). Ordinal logistic models use cumulative probabilities up to a threshold, thereby making the range of ordinal categories binary at that threshold (see link function below) (PennState, no date). Model

parameters are estimated by maximum likelihood; a method which identifies the strongest linear combination of independent variables, that increases the likelihood of detecting the dependent variable (Stoltzfus, 2011).

Logit link function: $\text{logit}(Y) = \log\left(\frac{p}{1-p}\right) = n(x)$

> Where $n(x)$ is the linear predictor. For logistic regression with a two category dependent variable (0 and 1) p is the probability that the dependent variable is 1. For ordinal logistic regression $p = p(Y \leq j)$ which is the cumulative probability that the dependent variable falls in category j or less, i.e. how likely the response is to be in category j or below versus in a category higher than j
>
> (Fox and Weisberg, 2011)

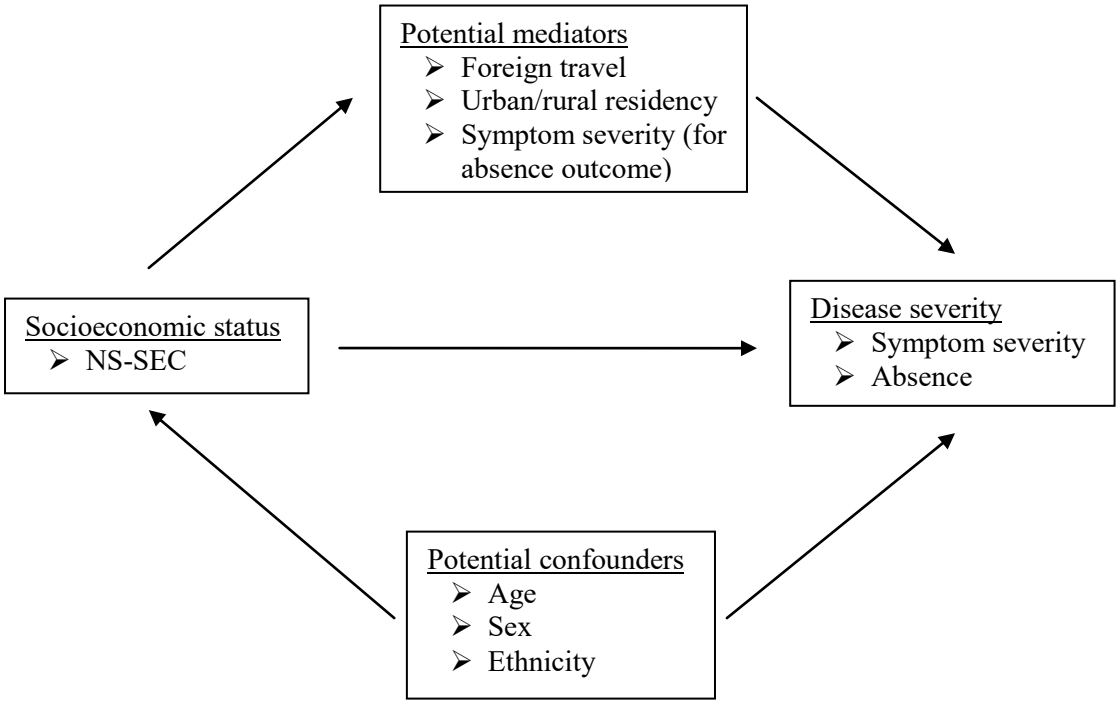
An assumption for both regression models is that the observations within a dataset are independent, i.e. there are no duplicate responses (Stoltzfus, 2011). Additionally, continuous independent variables should be linearly related to the log-odds of the dependent variable (Stoltzfus, 2011). For the continuous age variable this assumption was assessed visually using generalised additive models (GAMs), as described below. An additional assumption of ordinal logistic regression is that the coefficients describing the relationship between each pair of outcome categories are the same (the proportional odds assumption) (Ripley, 2017; UCLA: Statistical Consulting Group, no date-a). A graphical method was used to test this assumption (Harrell, 2001; UCLA: Statistical Consulting Group, no date-a).

GAMs were used to visually assess the linear relationship between the continuous age variable and the outcomes. A GAM is a type of GLM in which the linear predictor incorporates smooth functions of the predictor variables (Hastie and Tibshirani, 1990). GAMs can be plotted to visualise the shape of the relationship between the dependent and independent variables. Using GAMs, there appeared to be a linear relationship between age and the log-odds of sickness absence, therefore age was included as a continuous variable when modeling the absence outcome (see Appendix 5 for plots). The relationship between age and symptom severity was non-linear, therefore a categorical age group variable (with categories: 5–14; 15–24; 25–44; 45–64 and 65+ years) was included when modeling the symptom severity outcome. The boundaries of the categories were chosen to demonstrate the relationship between age and symptom severity as shown in Appendix 5.

A hierarchical approach was used for the multivariate regression modeling. Firstly, baseline models were fitted for each of the two outcomes (symptom severity and sickness absence)

with age, sex and ethnicity as independent variables. Secondly, NS-SEC was added as an additional independent variable to the models, and the improvement in model fit was tested using generalised likelihood ratio chi-square statistics to compare nested models. Thirdly, additional covariates (recent foreign travel and urban/rural residency) were included and tested to assess whether they improved the model fit. Finally, to explore whether differences in disease severity explained any association between SES and sickness absence, symptom severity was added as a control variable to the model with sickness absence as an outcome. A logic model, detailing the hypothetical causal pathway between SES and measures of IID severity, and the potential confounding or mediating effects of the covariates, is shown in Figure 3.1. This model only features the covariates that were available to analyse.

Figure 3.1 Logic model for IID2 analysis



An alpha level of 0.05 was used to define statistical significance. All analyses were conducted using R statistical software, version 3.3.1.

Sensitivity analyses

Several robustness tests were undertaken. The analyses were repeated using alternative cut-offs for the symptom severity score categories (mild 2–14; moderate 15–27; severe 28–40); implementing linear regression for the symptom severity score; including recurrent episodes

of IID within the same individual with clustering accounted for in mixed-effects models; including cases of all ages; and stratifying results by child and adult age groups. For the age stratified analyses, children were defined as aged <16 years, since this definition was used in the IID1 and IID2 studies (Tam et al., 2012b; Tam, Rodrigues and O'Brien, 2003).

Listwise deletion was used as the method of handling missing data. For the two outcomes, cases with missing data within any of the variables to be included in the models were excluded, to permit comparison of nested models with identical cases. Sensitivity analyses were performed using multiple imputation by chained equations to impute missing data values for all of the variables included in the models. The methods that were used are described in detail below.

Sensitivity analysis – multiple imputation

Listwise deletion is a common approach to deal with missing data, and is provided as the default option for analyses in most statistical software packages (Kang, 2013). It involves the exclusion of cases from the analysis if any values are missing within the variables of interest. However, listwise deletion can be problematic in that it reduces the sample size and thus statistical power of the analysis (Graham JW, 2009), and can produce biased estimates when missing data are not missing completely at random. Alternative methods such as multiple imputation can take into account the nature of the missing data (Dong and Peng, 2013), and preserve the sample size.

Multiple imputation generates *m* imputed datasets which contain estimated values for missing data points (Dong and Peng, 2013). The statistical modeling procedure (in this case ordinal logistic regression for the symptom severity outcome, and logistic regression for the sickness absence outcome) is then applied to each of the *m* imputed datasets and the resulting *m* parameter estimates are pooled to produce a single estimate with its standard error. The standard error incorporates the uncertainty due to the imputation of the missing data and the uncertainty due to the modeling procedure (Azur et al., 2011).

Multivariate imputation by chained equations (MICE) was used to impute the missing data, and this was achieved by using the 'mice' package version 2.25 in R (version 3.3.1). In MICE, regression models are used to estimate the missing data values within each variable, conditional on the other variables included in the model (Azur et al., 2011). This particular

method was chosen because each variable can be modelled according to its distribution and therefore categorical as well as continuous data can be imputed.

Nature of missing data

Within any one variable in a dataset, missing data can be either: missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR). These missing mechanisms have been described by Little and Rubin (2002):

- If the missing data does not depend on the observed values of a data matrix or the missing values (i.e. unobserved values), it is said to be MCAR
- If the missing data depend on the observed values but not the missing values, it is said to be MAR
- If the missing data depend on the missing values in the data matrix, it is said to be MNAR

An assumption of most multiple imputation methods is that the missing data mechanism can be ignored if the missingness is not related to the unobserved data (Siddique, Harel and Crespi, 2012; Rubin, 1976). Thus, standard multiple imputation methods can be used in situations where missing data are MCAR or MAR.

Before performing multiple imputation, it is therefore important to firstly consider whether the missing data mechanism could be MNAR. In this case, it seems unlikely that item non-response would be deliberate, i.e. not random. Participants were questioned about the occupation of the main-earner in their household as a means of deriving SES, rather than a more sensitive topic such as their income which can result in non-random missingness (Kim et al., 2007). The symptom severity variable was derived from nine separate IID symptom variables and again it seems unlikely that item non-response would be deliberate. An exception to this is the potential for deliberate item non-response due to a parent's inability to answer questions on the presence of subjective symptoms such as headache and nausea for their young child. However, the effects of this should have been minimised by only including cases $\geq$5 years of age in the analysis. Absence non-response could potentially be related to the participant not wishing to disclose that they were absent from work if they were absent, however again this seems unlikely given that the data were collected as part of a population-based survey, in settings outside of the participants' workplace environments.

Whilst multiple imputation can be performed in situations where missing data is either MCAR or MAR, it can be informative to distinguish between the two mechanisms. For example, where missing data are MCAR, methods to handle the missing data such as listwise deletion will theoretically produce unbiased estimates (Nakai and Weiming, 2011), and thus the need for multiple imputation may be reduced. To explore the nature of the missing data, analyses can be performed to assess whether the missingness in one variable can be predicted by any another variables. If the missingness in one variable can be explained by another variable, this supports the idea that the missing mechanism is MAR (Osborne, 2013). Univariate logistic regression analyses were therefore performed to investigate whether the missingness in the symptom severity, sickness absence and NS-SEC variables could have been explained by other variables in the dataset (Appendix 5).

Selection of variables to be included in multiple imputation model

Van Buuren and Groothuis-Oudshoorn (2011) set out three rules for selecting the most appropriate variables to include in a multiple imputation model. Firstly, all variables that will be used in the final model (after imputation) should be included in the imputation model. This includes the dependent variable. Omitting the dependent variable from the imputation model has been shown to bias the results towards the null when the dependent variable is related to the independent variables that are being imputed (Sterne et al., 2009; Moons et al., 2006). Secondly, the variables that are related to the missingness within the variables to be imputed should be included in the imputation model. This takes into account the nature of the missing data (Dong and Peng, 2013). Thirdly, variables that are predictors of the variables to be imputed should be included in the imputation model, to increase the precision of the estimates (Dong and Peng, 2013).

As mentioned, MICE allows each variable to be modelled according to its distribution. For continuous variables, predictive mean matching was specified as the imputation method. Predictive mean matching is a semi-parametric method and when used within the R mice package with untransformed skewed data, has been shown to produce precise estimates within 10% accuracy with up to 25% missingness (Van Buuren and Groothuis-Oudshoorn, 2011; Marshall et al., 2010). Logistic regression was used as the imputation method for binary variables, and a multinomial model was used for unordered factors of two or more levels (Van Buuren and Groothuis-Oudshoorn, 2011). Ordinal logistic regression was used to impute missing data within ordinal variables.

Running MICE and assessing convergence

To run the multiple imputation model the number of multiply imputed datasets required needs to be specified, as well as the number of iterations required. It has been suggested that the number of multiply imputed datasets needed is approximately similar to the percentage of cases that have any missing data (Allison, 2012; Bodner, 2008; Graham, Olchowski and Gilreath, 2007). For this analysis, 43.6% of cases had missing data on one or more variables, and therefore approximately 40 multiply imputed datasets were required.

Within an iteration cycle, each variable is visited and the missing values imputed using the imputation method specified (Azur et al., 2011). The number of iterations needed depends on how quickly the sampling distributions of the imputed values converge and become stable (Bouhlila and Sellaouti, 2013). For MICE convergence can be fast and therefore the number of iterations required can be small (Van Buuren and Groothuis-Oudshoorn, 2011), however for this analysis, due to the large amount of missing data 100 iterations were used.

After running the multiple imputation model, convergence was checked by plotting the mean and variance of the imputed values per iteration for each variable. Additionally to check the plausibility of the imputed values, the densities for the observed and imputed values were plotted.

Analysis with multiply imputed datasets

The dependent symptom severity score variable was included in the imputation model because omitting this variable implies that there is no relationship between the dependent and independent variables, resulting in biased estimates towards the null (Graham JW, 2009). However it has been suggested that including the imputed dependent variable values in the final analysis offers no advantages and can add unnecessary random variation to the estimates due to simulation error in the imputed dependent variable values (von Hippel, 2007). Therefore a method which involves imputing the dependent variable, but in the final analysis including only those cases that have observed dependent variable values, has been suggested as an alternative (von Hippel, 2007). It has been put forward that this method, named multiple imputation, then deletion (MID), provides less variable point estimates and more accurate standard error estimates compared to analyses that retain the imputed dependent variable values, especially when there are large amounts of missing data within the dependent variable (von Hippel, 2007). On the other hand, Young and Johnson (2010) found that the results produced by both methods were similar, using a dataset that contained missing data within the dependent and independent variables, the former having 11.4%

missing data. Both methods for dealing with imputed dependent variable data were used in this analysis for comparison.

Ethical considerations

This analysis was performed using anonymised datasets. The original IID2 study was granted ethical approval by the North West Research Ethics Committee (07/MRE08/5) on 19th April 2007 (Tam et al., 2012b). Written informed consent was obtained from all participants and also the parent or guardian of child participants. When entering the study, participants gave consent for their anonymised data to be used for future analyses.

3.3 STUDY 3

Study 3 was a cross-sectional ecological analysis of routinely collected HES data, which aimed to explore the relationship between neighbourhood income deprivation and emergency hospital admission rates and the duration of admissions for IID. The following section describes the data sources utilised, the study design that was employed to address the aims of the research, the selection of dependent and independent variables, and the methods used to analyse the data.

Data sources

Inpatient data obtained from HES were used to retrieve information on the outcomes of interest for the analysis. HES is a data warehouse, containing information on all A&E attendances, hospital admissions and outpatient appointments in NHS hospitals in England (NHS Digital, no date). The data are collected for administrative purposes (e.g. to calculate payments owed to hospitals for care that has been provided), however HES has also been designed for secondary data analysis (NHS Digital, no date).

Several steps are involved in the collection and processing of HES data. Firstly, healthcare providers collect certain administrative and clinical data whilst providing care for patients. For example, a patient's primary diagnosis and additional secondary/subsidiary diagnoses are recorded using ICD diagnosis codes. Data such as these are submitted to a secure data

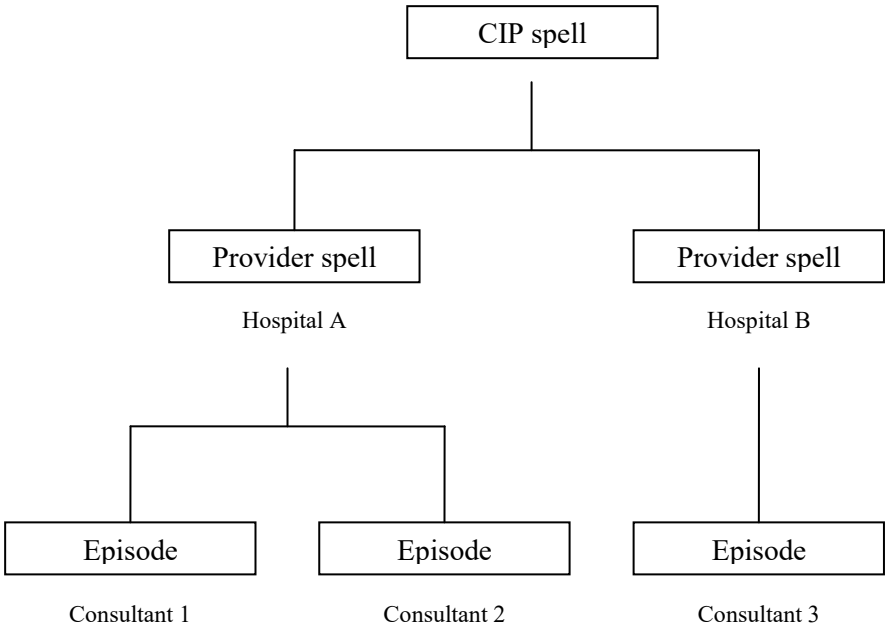
repository called the Secondary Uses Service (SUS), managed by NHS Digital (NHS Digital, 2016b). Each month, raw data are extracted from SUS and sent to HES, where the data are cleaned, duplicates removed and additional fields are derived (NHS Digital, 2016b). For example, using a patient's postcode, several additional fields can be derived relating to their area of residence, such as the LSOA and IMD score. Where data are missing, incorrectly submitted and when duplicates are found, the HES Data Quality team liaise with healthcare providers to correct inaccuracies (NHS Digital, 2016b).

Following the cleaning process, HES data are ready to be published. A number of standard analyses are published online by NHS Digital. For this analysis, data were obtained from the Integrated Longitudinal Research Resource (ILRR) at the University of Liverpool. The ILRR is provided with a pseudo-anonymised extract of HES admitted patient care data via a data sharing agreement with NHS Digital. NHS Digital pseudo-anonymises HES data by assigning a unique but anonymous identification number to each patient (Health and Social Care Information Centre [HSCIC], 2015). For this study, a data access request was made to the ILRR to access anonymised and aggregated data, as detailed below.

When using HES data, a patient's stay in hospital is described as a spell. There are two types of spells: Provider, and Continuous Inpatient (CIP). A CIP spell is a period of continuous care within the NHS for a patient, regardless of any hospital transfers that have taken place (HSCIC, 2014). When a patient is admitted a CIP spell starts, and when a patient is discharged or dies the CIP spell ends. A CIP spell can contain one or more Provider spells, which are periods of care that have taken place in one hospital (HSCIC, 2014). If an admitted patient is discharged from one hospital and transferred to a different hospital, one Provider spell ends and a new Provider spell starts. Spells are also made up of one or more episodes, which are periods of care under one consultant (HSCIC, 2014). For the majority of patients, their CIP spell has only one Provider spell and one episode (HSCIC, 2014).

This analysis was performed using Provider spells as measures of admissions, since the data necessary to calculate CIP spells from Provider spells was not available in the data excerpt obtained from NHS Digital. It was assumed that the use of Provider spells instead of CIP spells would have had a minimal impact on the results based on previous literature (Busby, Purdy and Hollingworth, 2017a), however this potential limitation is evaluated further in the discussion. The hierarchy between spells and episodes is demonstrated in Figure 3.2, for a fictional patient with three episodes.

Figure 3.2 Example of relationship between spells and episodes in HES



Source: Adapted from HSCIC (2014)

For this analysis, Provider spells (which I will refer to as admissions) and data on admission days for IID were extracted from a pseudo-anonymised HES data excerpt by an ILRR data scientist at the University of Liverpool, and were aggregated to the LSOA of residence of the patient. Admission data were also aggregated into three age groups (children aged 0–14 years; adults aged 15–64 years; adults aged 65+ years) for each LSOA. Following the suppression/removal of cells with small counts (n<5) that also had small underlying population sizes (n<1000), the dataset was released to me. To conduct the analysis, I linked the HES dataset to various openly available national datasets, such as 2011 Census data obtained from the ONS Nomis website, mid-year population estimates published by the ONS, data published by the Consumer Data Research Centre (CDRC), and indices of deprivation data published by the Department for Communities and Local Government. The specific data used in the analysis are discussed subsequently.

Population

The population under investigation was the population of England. English neighbourhoods were the units of analysis in this study, and these were defined using LSOA geographical boundaries. LSOAs and the larger sized Middle-Layer Super Output Areas (MSOAs) were initially introduced following the 2001 Census, and they have since become the standard

units for presenting local statistics (ONS, 2016). LSOAs have minimum and maximum household and population thresholds; each containing 1000 to 3000 people (ONS, 2016). In England, there were 32,482 LSOAs following the 2001 Census, and this number increased to 32,844 LSOAs following the 2011 Census (ONS, 2016). This analysis was based on English LSOAs derived from the 2011 Census.

Study design

A cross-sectional ecological analysis was performed using data aggregated at the LSOA level, to explore the relationship between neighbourhood deprivation and emergency hospitalisations for IID, over a seven year period from 1st January 2009 to 31st December 2015.

Outcomes

Outcome 1: Emergency hospital admission rates for IID

The first outcome of interest was rates of emergency hospital inpatient stays (Provider spells) with IID as the primary diagnosis. A diagnosis of IID was defined using ICD-10 codes: A00–A09 (intestinal infectious diseases), or K52.9 (unspecified non-infective gastroenteritis and colitis). The code K52.9 was included in the definition, since unspecified acute gastroenteritis cases were regularly coded within the NHS as digestive conditions (K52.9) until 2010–11, however from 2011–12 cases started to be coded as infectious conditions (NHS England, 2014). Previous research recommends including both A09 and K52.9 codes, to capture acute gastroenteritis of undetermined aetiology, especially if the study period includes the year 2009 (Wilson, Deeks and Rosella, 2015). Age specific emergency hospital admission rates were calculated per LSOA over the seven year period from 2009–15, using mid-year population estimates published by the ONS. The data extract available to me covered this seven year period only, and emergency admissions were analysed because they make up the majority of admissions for IID (as shown in Table 2.4).

Outcome 2: Admission days per emergency admission for IID

The second outcome of interest was hospital admission duration (length of stay) for IID-related emergency admissions. This was calculated as the total number of hospital admission days divided by the total number of emergency hospital admissions for IID, over the period

2009–15, per LSOA. Data were also aggregated into three age groups per LSOA. Emergency admissions for IID were defined as above. LSOAs that had zero emergency hospital admissions for IID over the period 2009–15 were excluded when analysing the admission days per admission outcome, since these LSOAs were considered not applicable to the analysis of the duration of stay.

Exposures

Primary exposure: Income deprivation

The primary exposure of interest was the income deprivation domain for each LSOA, used to derive the 2015 English IMD scores. The overall IMD score is a relative measure of deprivation provided at the LSOA level, and is derived from seven domains of deprivation relating to income, employment, education, health, crime, the living environment and barriers to housing and services (Department for Communities and Local Government, 2015). The income deprivation domain accounts for 22.5% of the overall IMD score, and measures the proportion of the population experiencing deprivation relating to low income. The definition of low income includes those who are out-of-work and those who are in work but have low earnings.

The income deprivation domain was chosen as the primary exposure of interest, over the IMD overall score, because the overall score contains a health domain and thus captures a measure of morbidity and disability which could be correlated with the health outcomes of interest. An additional advantage of using the income deprivation domain compared to the overall IMD and the education, health, crime, living environment and barriers to housing and services domains, was that the scores for the income deprivation domain could be used to compare areas on an absolute scale. For example, if an area has a score of 0.38 in the income deprivation domain, this means that 38% of the population is income deprived in that area (Smith et al., 2015). Using the income domain over the other domains therefore meant that the results were easier to interpret and were more meaningful. Furthermore, other domains, such as those describing the living environment, contain measures that may have been more difficult to make sense of in relation to IID admission rates, such as the proportion of homes with central heating and number of road traffic accidents. The income deprivation domain from the IMD 2015 scores were used, since these scores were based on LSOAs from the 2011 Census and most of the data used to derive the IMD 2015 scores relates to the tax year 2012–13 (Department for Communities and Local Government, 2015).

Covariates: Age, ethnicity, comorbidity, ruality, distance to GP and hospital

The analysis was stratified by three age groups (children aged 0–14 years; adults aged 15–64 years and adults aged 65+ years). This was to control for the potential confounding effect of age on the relationship between deprivation and the hospitalisation outcomes for IID, as demonstrated by previous research (Olowokure et al., 1999). Data on the age structure of each LSOA was derived using age specific mid-year population estimates published by the ONS.

Additionally, several neighbourhood level covariates were included in the analysis. Data from the 2011 Census were used to derive variables relating to ethnicity and the prevalence of long-term health problems per LSOA. I calculated the proportion of the population that were of White ethnicity by dividing the number of individuals who defined their ethnicity as White in the 2011 Census by the usual resident population (also derived from the 2011 Census). The proportion of the population that reported having a long-term health problem or disability was calculated in a similar fashion. The 2011 Census defines a long-term health problem or disability as that which limits a person's day-to-day activities, and has lasted or is expected to last at least 12 months, including problems that are related to old age (ONS, 2013). Since age specific data were available, I calculated the ethnicity and long-term health problem variables for three separate age groups (children aged 0–14 years; adults aged 15–64 years; adults aged 65+ years) per LSOA. Whilst there were no missing data within these variables, some observations had been swapped or imputed by the data owners to protect against the disclosure of personal information (ONS, 2012). Swapped households are matched on basic characteristics to preserve data quality, and most swapping is done within the same MSOA (ONS, 2012).

Additionally, the rural/urban classification of each LSOA obtained from the ONS, was included as a covariate. The rural/urban classification is calculated at the smallest available geography (the output area [OA] level), and larger geographies such as LSOAs, are assigned the category that the majority of their constituent OAs have been assigned (Bibby and Brindley, 2013). OAs are classified as ‘urban’ if they are allocated to a 2011 built-up area with a population of 10,000 people or more, while all remaining OAs are classified as ‘rural’ (Bibby and Brindley, 2013).

Finally, variables indicating the average network (road) distance in kilometres to the nearest GP, and nearest hospital with an A&E department, per LSOA were included in the analysis. These variables were obtained from colleagues at the University of Liverpool, however the

data have also been published on the CDRC website (CDRC, 2017). The network distance in kilometres was calculated by deriving the fastest route by car to get from each postcode to the nearest GP and nearest hospital with an A&E department. The distances for each postcode within an LSOA were averaged to provide data at the LSOA level.

Age and ethnicity were considered to be potential confounders of the relationship between deprivation and hospital admission and admission duration for IID, as they may have been related to both deprivation and the hospitalisation outcomes, but were unlikely to be in the hypothetical causal pathway between them since deprivation does not determine age or ethnicity (Figure 3.3). It was considered that the rural/urban classification, the distance to health service variables, and the proportion of the population with a long-term health problem or disability, could have been related to deprivation and the hospitalisation outcomes, and also may have featured in the hypothetical causal pathway between them (Figure 3.3). For example, levels of deprivation are generally lower in rural compared to urban areas (Gartner et al., 2008), and rural residency may put individuals at increased exposure to certain pathogens such as STEC which can result in severe sequalae (Byrne et al., 2015). Relatedly, perhaps for employment reasons, individuals of lower SES tend to live closer to city centres (Cuberes and Roberts, 2015), which might mean they live in closer proximity to health services such as GPs and hospitals with A&E departments. Previous studies have found associations between shorter distances from hospital and increased emergency hospital admission rates (Busby, Purdy and Hollingworth, 2017b; Bankart et al., 2011), however whether the distance from an individual's area of residence to hospital is associated with length of stay is less clear (Heys, Rajan and Blair, 2017; Agboado, Peters and Donkin, 2012). Anecdotally, individuals who live closer to their GP may access primary care services more readily and/or at an earlier stage of IID progression, which may have an impact on hospital admissions and admission duration for IID.

Statistical analysis

To investigate inequalities in emergency hospital admission rates for IID, negative binomial regression models were used. Negative binomial models are a type of GLM usually used to model over-dispersed count data, whereby the conditional variance exceeds the conditional mean (Grootendorst, 2002; Gardner, Mulvey and Shaw, 1995). Count data are typically whole numbers (integers) that are greater than or equal to zero; such as the number of hospital admissions per LSOA. Negative binomial models can be thought of as an extension of the Poisson regression model. Poisson regression is suitable for modeling Poisson

distributed count dependent variables, whereby the distribution of counts have a variance that is equal to the mean (Grootendorst, 2002). If the distribution of counts has a variance that exceeds the mean, a Poisson distribution may not be an appropriate model. Instead, a negative binomial distribution may be more appropriate, since it has an additional parameter to the Poisson (the dispersion parameter) to model the over-dispersion (Grootendorst, 2002; Gardner, Mulvey and Shaw, 1995).

As mentioned previously in the methods for Study 2 of this thesis, GLMs incorporate a link function to translate the scale of the mean of the dependent variable to the scale of the linear predictor (Fox and Weisberg, 2011). Negative binomial and Poisson regression models typically use a log link function. For this analysis, counts of emergency hospital admissions per LSOA were modelled with the log of the population per LSOA as an 'offset' variable, indicating the maximum number of hospital admissions that could have occurred per LSOA (see below). The log of the population was used since the negative binomial model uses a log link function (Zwilling, 2013). Model parameters are estimated by maximum likelihood; a method which identifies the strongest linear combination of independent variables, that increases the likelihood of detecting the dependent variable (Stoltzfus, 2011).

Model for rates using offset: $\log \frac{\mu}{t} = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$

which is equivalent to: $\log \mu = \log t + \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$

Where μ is the expected count of hospital admissions for IID per LSOA, and t is the population per LSOA ($\log t$ is the offset)

As mentioned, the negative binomial model assumes the conditional means are not equal to the conditional variances, and as such a Poisson distribution may be more appropriate if the conditional means and variances are equal. To test whether a negative binomial model was appropriate for this analysis, the data were modeling using both Poisson and negative binomial regression, and the fit of the models to the data were compared using the likelihood ratio test (UCLA: Statistical Consulting Group, no date-b). An additional assumption is that the observations within a dataset should be independent (Fox and Weisberg, 2011).

The second outcome of interest, admission days per emergency hospital admission for IID, was modelled using linear regression. Linear regression can be used to model normally distributed continuous dependent variables. It incorporates an identity link function, wherein the mean of the dependent variable is modelled directly (Fox and Weisberg, 2011). Linear

regression was chosen to model the admission days per admission outcome, over Poisson or negative binomial regression, because the admission days variable was not a discrete count outcome. The admission days per admission dependent variable was positively skewed, and therefore to improve the model fit, the dependent variable was log transformed to create a more normal distribution for modeling.

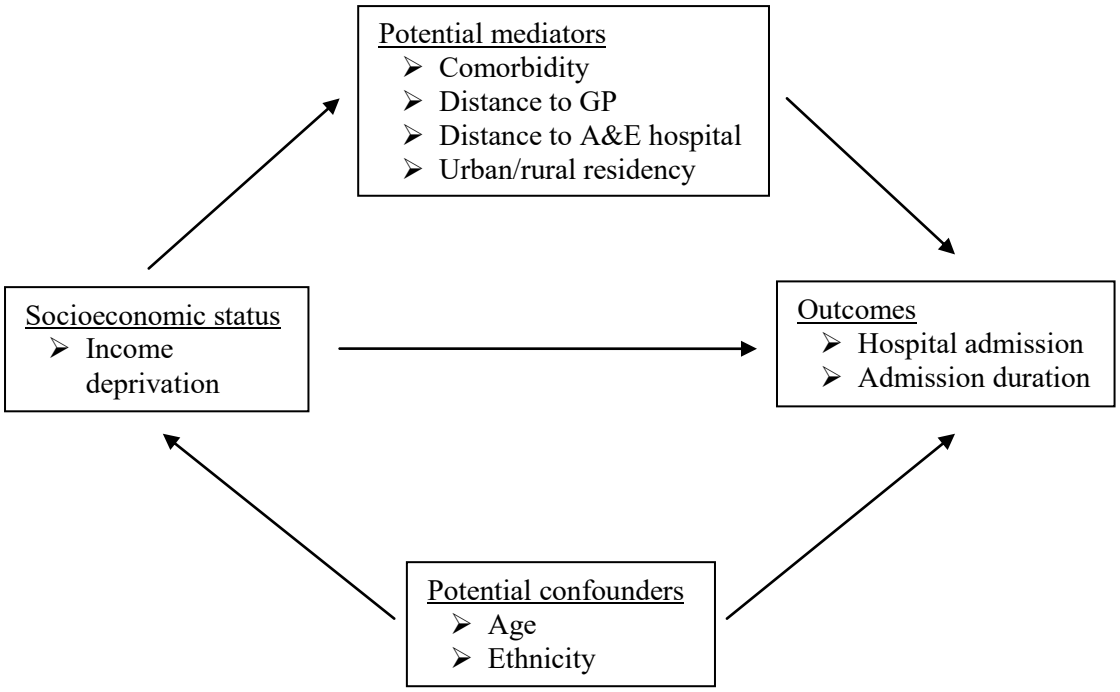
There are several key assumptions of linear regression, in addition to the assumption of independence of observations, as mentioned above. Firstly, the differences between the observed and predicted dependent variable values (the residuals), are assumed to be normally distributed (Weisberg, 2005). This was assessed by plotting a histogram of the residuals. Additionally, the variance of the error terms should be similar for all values of the independent variables. Equality of variance (homoscedasticity) was checked by plotting the residuals against the fitted values. Furthermore, a linear relationship between the dependent and independent variables is assumed to exist, which was assessed visually using GAMs.

In exploratory analysis, GAMs were used to visually assess the relationships between the continuous independent variables and the scaled risk of the hospitalisation outcomes. All of the continuous independent variables were approximately linearly related to the admission duration outcome, and were therefore retained as continuous variables when modeling this outcome. On the other hand, the relationships between the scaled rate of emergency hospital admissions and income deprivation, ethnicity and long-term health problems, for some age groups, appeared to be non-linear (Appendix 6). Therefore these variables were included as categorical variables when modeling the emergency hospital admission rate outcome. Income deprivation was categorised as quintiles, since the IMD (of which income deprivation is a domain) is commonly operationalised as quintiles. The average percentage of those of White ethnicity was relatively similar for children, adults and older adults, and therefore to aid interpretation the ethnicity variable was categorised with the boundaries: $\leq$70%; >70% to $\leq$90%; >90% for all age groups. In contrast, the average percentage of those with long-term health problems varied markedly across the age groups, and therefore it was more appropriate to categorise this variable as quartiles for each separate age group. Similar results to those presented in the main analysis were observed when the ethnicity variable was operationalised as tertiles, and when the long-term health problem variable was operationalised as quintiles (data not shown).

For both outcomes of interest (emergency hospital admissions and admission days per admission), a hierarchical approach was used for the multivariate regression modeling. The analysis was stratified by three age groups, to control for the potential confounding effects of

age on the relationship between income deprivation and the hospitalisation outcomes. Baseline models were fitted for each of the two outcomes with ethnicity as an independent variable. Income deprivation was then added as an additional independent variable to the models. Additional covariates, such as the proportion of the population with a long-term health problem or disability, and the geographical variables (rurality and distances to health services) were included to assess whether they explained any association between income deprivation and the hospitalisation outcomes. A logic model, detailing the hypothetical causal pathway between income deprivation and the hospitalisation outcomes, and the potential effects of the covariates, is shown in Figure 3.3. This model only features the covariates that were available to analyse, for example sex was not included as admission data were not available by sex.

Figure 3.3 Logic model for HES analysis



Results from the negative binomial and linear regression models were expressed as incident rate ratios (IRRs) and regression coefficients, respectively. The exponentiated regression coefficients were used to calculate the percent change (expressed as the percent increase or decrease in the outcome variable for every unit increase in the exposure variable). The income deprivation, long-term health problem and ethnicity exposure variables were entered into the regression models for the admission duration outcome, in units of 10% points.

Therefore the percent change was calculated for every 10% point increase in these exposure variables.

An alpha level of 0.05 was used to define statistical significance. All analyses were conducted using R statistical software, version 3.3.1.

Sensitivity analyses

Robustness tests were performed to investigate the potential impact of including the ICD-10 code K52.9 (unspecified non-infective gastroenteritis and colitis) in the definition of an emergency hospital admission for IID. The analyses were repeated using a different definition of IID, that excluded codes K52.9 and A09.9 (gastroenteritis and colitis of unspecified origin). This definition of IID was more specific, but was probably less sensitive, given that possible cases of IID were likely excluded.

Ethical considerations

For this analysis, a Data Access Request to the ILRR at the University of Liverpool was made to access anonymised and aggregated HES datasets. Total counts of admissions and total admission days for IID, aggregated over a seven year period by LSOA and three age groups were requested. Cells with small counts ($n<5$) that also had small underlying pooled population sizes ($n<1000$) were suppressed as required by the Anonymisation Standard for Publishing Health and Social Care Data Specification (Information Standards Board for Health and Social Care, 2013) before being released to me. This minimised any risk of re-identification.

Chapter 4

Results: Study 1

Relationship between SES and GI infections in high income countries: a systematic review and meta-analysis

4.1 ABSTRACT

Background

Several studies have measured the association between SES and the incidence of GI infections in high income countries. Yet the direction and magnitude of this association remain unclear. A number of studies have observed an increased risk of GI infection amongst those of lower SES, and others have observed the opposite; an increased risk of infection amongst those of higher SES. Therefore a systematic review was conducted to make sense of these contradictory findings, and explore possible sources of heterogeneity in effect estimates reported in the literature.

Methods

MEDLINE (Ovid), Scopus, Web of Science Core Collection and grey literature were searched from 1980 to October 2015 for observational studies reporting a quantitative association between SES and the incidence of GI infections, in a representative population sample from a member-country of the OECD. Study quality was assessed using the LQAT. Harvest plots, meta-regression and subgroup meta-analyses were used to synthesise study findings and explore potential sources of heterogeneity (identified a priori), such as pathogen type, country, age of the participants, measure of SES and the source of GI infection cases.

Results

In total, 6049 studies were identified, and 102 met the inclusion criteria. In multivariate meta-regression analysis, age was identified as the only statistically significant potential effect modifier of the relationship between SES and GI infection risk. For children, risk of GI infection was statistically significantly higher for those of lower SES versus high (RR 1.51; 95% CI 1.26–1.83), but there was no significant difference for adults (RR 0.79; 95% CI 0.58–1.06). For studies that identified cases via hospitals the pooled risk ratio was 1.47 (95% CI 1.19–1.82), compared to population-based surveys 1.07 (95% CI 0.88–1.29), however the majority of studies that analysed hospitalised cases also analysed children.

Conclusions

Disadvantaged children, but not adults, appear to have greater risk of GI infection compared to their more advantaged counterparts. Overall age explained a very small proportion of the heterogeneity observed across the studies as a whole. Further research is needed to better understand inequalities in the risk of GI infection requiring hospitalization, especially in relation to whether the risk differs between adults and children.

4.2 INTRODUCTION

Examining the extent of socioeconomic inequalities in the risk of GI infections may offer important insights that enhance understanding of inequalities in the consequences of GI infections. This is because the interpretation of studies that have examined inequalities in consequences such as healthcare utilisation for GI infections in the general population, are hindered by uncertainties about the social patterning of the incidence of infection. As highlighted in Chapter 2, studies that have examined the risk of primary and secondary care presentation for GI infections in the general population have tended to find an increased risk for those of lower SES compared to high. This could reflect increased disease severity or a propensity for healthcare-seeking amongst those of lower SES. Alternatively, the results could be a simple reflection of increased disease incidence amongst those of lower SES compared to high. In other words, individuals of lower SES may be more likely to access healthcare services because they have a greater need for such services. The current study explores this latter explanation. Assessing and gaining a better understanding of the relative importance of these possible explanations may help to inform the development of effective solutions to address any inequalities observed in the consequences of GI infections.

As mentioned in Chapter 2, several studies have measured the association between SES and the incidence of GI infections in high income countries. Yet the direction and magnitude of this association remain unclear. A number of studies have observed an increased risk of GI infection amongst more socioeconomically disadvantaged groups (Beale et al., 2010; Özkan et al., 2007; Ludvigsson et al., 2006; Etiler, Velipasaoglu and Aktekin, 2004; Bozkurt, Özgür and Özçirpici, 2003; Bozkurt, Özgür and Özçirpici, 1999; Baker, Taylor and Henderson, 1998; Turkish Ministry of Health, 1995; Eaton-Evans and Dugdale, 1987). Whilst others have observed the opposite; an increased risk of infection amongst more socioeconomically advantaged groups (Adams et al., 2017; Pollard et al., 2014; Van Cauteren et al., 2012; Scallan et al., 2004; Herikstad et al., 2002; De Wit et al., 2001a; Fein, Lin and Levy, 1995).

There are a number of potential explanations for the contrasting findings observed in the literature. For example, studies that have measured this association have been conducted in different countries, such as the UK, France, the USA, Australia and Turkey. These countries differ in several ways including economic development and climate; factors which may or may not modify the relationship between SES and the incidence of GI infections. Additionally, studies have used different measures of SES, from individually-based measured such as education level, to area-based measures such as the IMD. Studies have

also utilised different study designs and controlled for a variety of potential confounding factors. Thus, there are many sources of clinical and methodological heterogeneity amongst studies that have measured the association between SES and the incidence of GI infections, however it is unknown to what extent these sources contribute to the contrasting results observed.

Additionally, comparing the results of studies that have sampled laboratory confirmed GI infection cases to studies that have sampled cases from population-based surveys may offer interesting insights. The potential modifying role of pathogen type on the relationship between SES and incidence of infection can be explored using studies that have sampled laboratory confirmed cases. However, laboratory notified cases will have consulted a healthcare professional for their illness and provided a stool specimen, and it is not known to what extent these cases differ, in terms of SES, to community cases. Cases identified via laboratories also represent only a small fraction of cases occurring in the community, and population-based surveys have greater scope for capturing the true burden of GI infections. Moreover, comparing the results of population-based surveys, to the results of studies that have sampled cases presenting to their GP or hospital, may offer clues as to whether individuals of lower SES have a greater need for healthcare services because they have a greater baseline risk of infection.

A systematic review is therefore warranted to summarise, organise and make sense of the contradictory findings observed in the literature, and to enhance understanding by comparing inequalities in the risk of infection amongst studies that have analysed GI infection cases occurring in the community and presenting to healthcare services.

4.3 AIM AND OBJECTIVES

Aim

- To systematically review current evidence on the relationship between SES and the incidence of symptomatic GI infections in high income countries, using studies that have identified cases via healthcare records, laboratory notifications and population-based surveys.

Objectives

- To assess the magnitude, statistical significance and direction of the association between SES and GI infection risk
- To evaluate possible sources of heterogeneity in effect estimates reported in the literature
- To investigate potential sources of bias in studies measuring the association
- To identify gaps in the knowledge base and areas for further research

4.4 METHODS

The methods are described in detail in Chapter 3.

Ethical considerations

This was a secondary piece of research and therefore did not require formal ethical approval. Despite this, there was an ethical obligation to conduct the review as objectively as possible, taking measures to minimise bias and critique research fairly (Hart, 1998).

4.5 RESULTS

Search results

- **Database search results**

The results from the database searches are displayed in Table 4.1. In total, 13,519 hits were retrieved from the three databases combined.

In MEDLINE, there was no available document filter and after limiting the results to studies that involved human subjects, that were published in English since 1980, a total of 1651 hits remained. In Scopus, the results were limited to the English language and publications that

had ‘human’ or ‘humans’ as keywords. Additionally, 1046 reviews, 64 editorials and 94 notes, and 178 items published between the years 1966 and 1979 were excluded.

In Web of Science Core Collection, there was no human subject filter, and so the results were refined to the English language. In total, 14 items described as editorial materials, 282 reviews, 9 notes, 4 meeting abstracts and 1 biographical item were excluded, and there were no references remaining published before the year 1980.

Table 4.1 Results of database search

		SEARCH TERMS							
		#1 SES terms	#2 GI infection terms	#3 OECD countries	#1 AND #2 AND #3	FILTERS (where available) →			
Database	**Date**	**NUMBER OF HITS**				**English**	**Human**	**Document**	**Published since 1980**
MEDLINE	13/10/15	1,351,793	570,945	2,591,382	2315	1878	1769	N/A	1651
Scopus	13/10/15	3,893,927	862,812	7,407,396	8061	7109	5830	4626	4448
Web of Science	13/10/15	2,272,138	611,880	3,852,328	3143	2953	N/A	2643	2643
				TOTAL	13,519				

After applying the database filters there were 8742 studies available for screening from the three databases combined, of which 2721 were identified as duplicates and removed (Figure 4.1).

Grey literature search results

The Google Scholar search, from 1st January 1980 to 31st December 2015, produced approximately 38,300 results. The first 100 studies in order of relevance were selected, and of these three studies were selected for full-text screening.

The Google search over the same time period retrieved approximately 579,000 results. Again, the first 100 hits in order of relevance were selected, and of these two studies were

selected for full-text screening. Of the five studies selected for full-text screening from the grey literature search, none were included in the final review.

Reference list search results

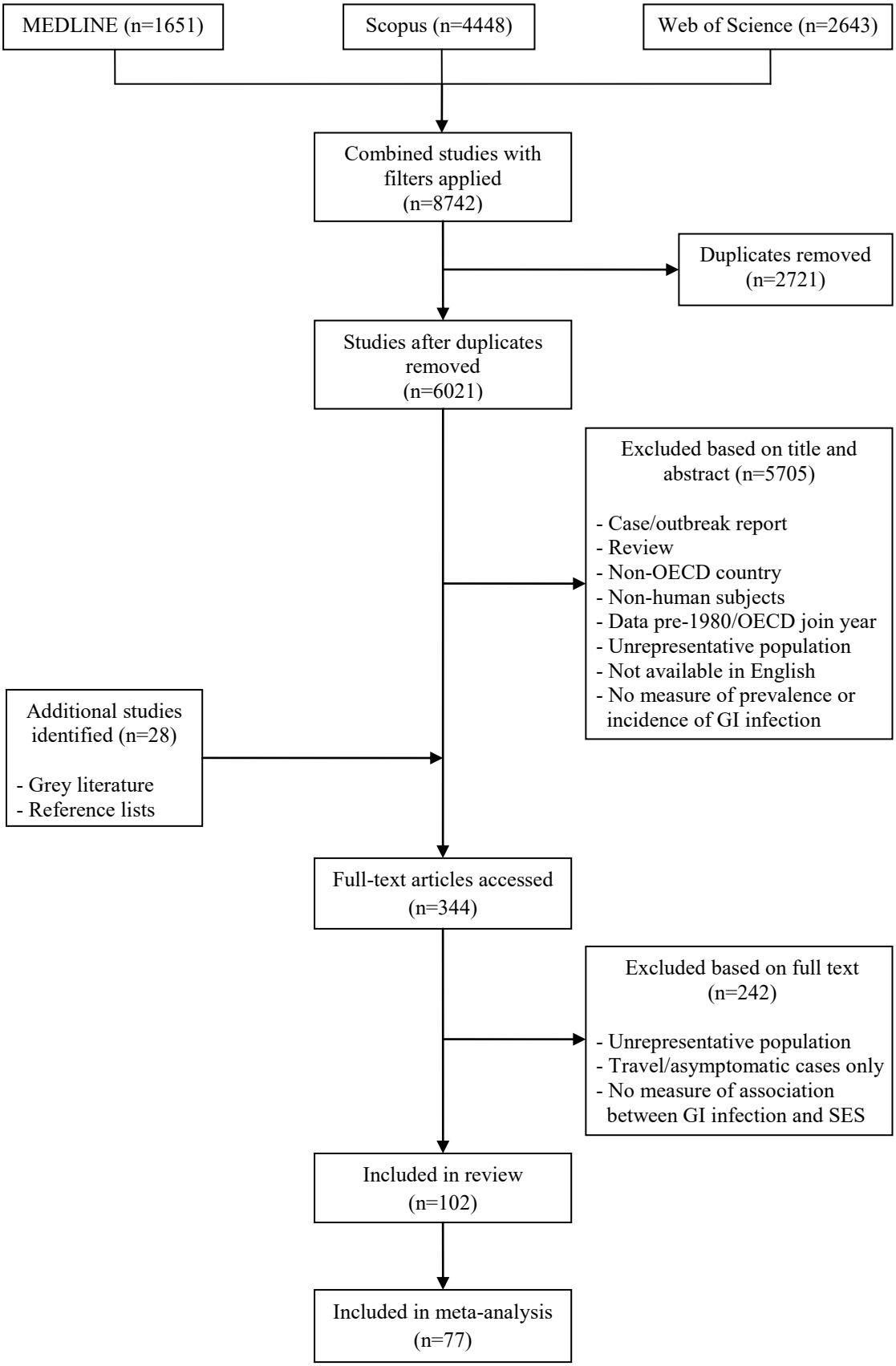
Hand searching of the reference lists of studies selected for inclusion in the review, resulted in the identification of 23 studies which were full-text screened. Of these, 16 were included in the final review.

The overall selection process is shown in Figure 4.1. Title and abstract screening was performed on 6021 studies, and those which mentioned assessment of the prevalence, incidence or risk of a GI infection of interest, using a general population sample were selected for full-text screening. Non-observational studies, case reports, outbreak reports, studies which did not use human subjects, had unrepresentative population samples, those that did not use subjects from populations of OECD countries, collected data before 1980, were not available in English or did not measure the prevalence or incidence of a GI infection of interest were excluded (n=5705).

Overall, 242 studies out of 344 were excluded based on full-text screening. The reasons for exclusion included: the use of an unrepresentative population sample (such as pregnant women, military personnel or those in institutions such as prisons); analysis of travel related or asymptomatic cases only; and no measurement of the association between GI infection risk and SES.

Where more than one study analysed the same cases using the same outcomes and exposures, only one study was included based on the study with the greatest focus and amount of information on the relationship between SES and GI infection risk. Majowicz, Horrocks and Bocking (2007) analysed data from two population-based telephone surveys, the first conducted in Hamilton, Ontario, Canada in 2001–2 and the second in British Columbia in 2002–3. Majowicz et al. (2004) analysed data from the first telephone survey only, using the same outcome and exposures as Majowicz, Horrocks and Bocking (2007), and so Majowicz et al. (2004) was excluded from the review.

Figure 4.1 Flowchart detailing selection of studies for inclusion



Britton et al. (2010) and Snel et al. (2009) analysed cryptosporidiosis and giardiasis infections in New Zealand, for the period from 1997 to 2006. Since both studies also used deprivation as the SES measure, it was decided that only one study could be included in the review. Snel et al. (2009) provided unadjusted rate ratios for five ordered categories of deprivation, whereas Britton et al. (2010) provided a rate ratio comparing two categories of deprivation adjusted for rurality and climate. The adjusted estimate provided by Britton et al. (2010) was considered more informative than the unadjusted estimates provided by Snel et al. (2009) and therefore the latter was excluded from the review. The results from both studies showed similar trends.

Finally, in total 102 studies were selected for inclusion in the review and 77 were eligible for inclusion in the meta-analysis (Figure 4.1) (see Appendix 4 for bibliography of included studies).

Study characteristics

The summary characteristics of the included studies are shown in Table 4.2. The majority of studies were conducted in Europe, had ecological study designs and used laboratory records to identify GI infection cases. Education level was identified as the most commonly used measure of SES across the studies. Most studies sampled participants of all ages, however 27 sampled children only, and eight sampled adults. Among the countries of origin of the included studies, Turkey was the only country classified as having high human development according to the Human Development Index, whereas all other countries had very high human development. Full details of the studies can be found in Appendix 4.

The majority of the studies were graded as low quality (n=56). Of these there were four cross-sectional, 35 ecological, eight cohort and nine case-control studies. Twenty-seven studies were graded as being of medium quality, including seven cross-sectional, four ecological, four cohort and 12 case-control studies. Finally, 19 studies were graded as high quality, seven cross-sectional, four ecological, four cohort and four case-control studies.

For the majority of studies, the investigation of the relationship between SES and the risk of GI infection was not a primary aim. Many studies analysed a dichotomous SES measure and did not investigate a social gradient in the risk of infection.

Table 4.2 Characteristics of included studies

Study Characteristics	Number of Studies
Total	102
Year of Publication	
Before 2000	17
2000–2005	15
2006–2010	38
After 2010	32
Level of Analysis	
Individual	59
Area	43
Region	
Asia	3
Europe	49
North America	34
Oceania	16
Study Design	
Case-control	25
Cohort	16
Cross-sectional	18
Ecological	43
Sample Size	
<200	3
200–1000	25
1001–5000	15
5001–10,000	9
10,001–100,000	5
>100,000	45
Quality	
High	19
Medium	27
Low	56
Age of Participants Sampled	
Children (<18 years old)	27
Adults	8
Mixed	61
Not stated	6
GI Infection Outcome	
Acute GI infection (syndromic)	41
Campylobacteriosis	20
Cryptosporidiosis	4
Giardiasis	3
Hepatitis A	3
Listeriosis	1
Norovirus	1
Rotavirus	3
Salmonellosis	8
Shigellosis	3
STEC	4
Yersiniosis	1
Multiple pathogens	10
SES Measure	
Deprivation	17
Education	22
Employment	7
Income	10
Occupation	8
Social class	10
Multiple measures	28

Method of Sampling Cases	
Population-based survey	30
GP presentations	5
Hospital admissions	13
Laboratory records	52
Multiple measures	2

See Appendix 4 for details of included studies
GI = Gastrointestinal; SES = Socioeconomic status; STEC = Shiga toxin-producing *E. coli*

Data analysis and synthesis

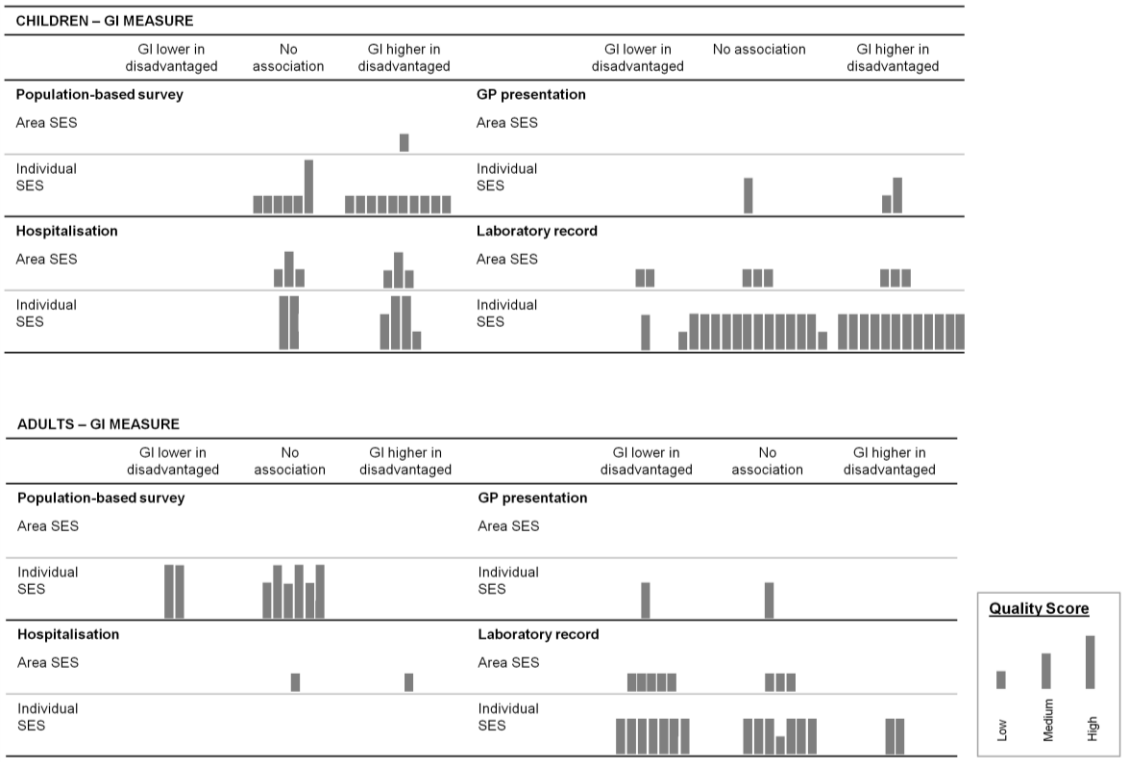
Harvest plots

To aid interpretation, point estimates for the association between SES and GI infection risk, calculated using either children only, or adults only, are presented in the harvest plots (Figures 4.2 and 4.3). Of the 102 studies included, there were 103 point estimates for adults or children specifically, and these estimates are represented graphically as bars in the harvest plot. Figure 4.2 shows the harvest plot for GI infection risk by SES, stratified by age, method of sampling GI infection cases, and SES measure. An additional harvest plot stratified by age, pathogen transmission route and SES measure is presented in Figure 4.3, displaying point estimates based on cases with a laboratory report.

The harvest plot displayed in Figure 4.2, illustrates that the relationship between SES and GI infection risk varies with age. The point estimates for children are shown on the upper half of the plot and the estimates for adults are shown on the lower half. For children, the majority of point estimates showed a higher risk of GI infection among disadvantaged children or no association between GI infection risk and SES; although several studies were of low quality as indicated by the height of the bars. The pattern for adults differed from that for children, with most point estimates weighting towards lower risk of GI infection among disadvantaged adults or no association.

There were no point estimates suggesting a lower risk of GI infection among disadvantaged children when cases were sampled via population-based surveys, GP presentation or hospital admissions (Figure 4.2). The vast majority of point estimates calculated using hospital admission data were based on children (86% [n=12/14]). Overall, a small number of point estimates were based on cases identified via GP presentation.

Figure 4.2 Harvest plot for risk of GI infection by SES, stratified by method used to sample cases, SES measure and age

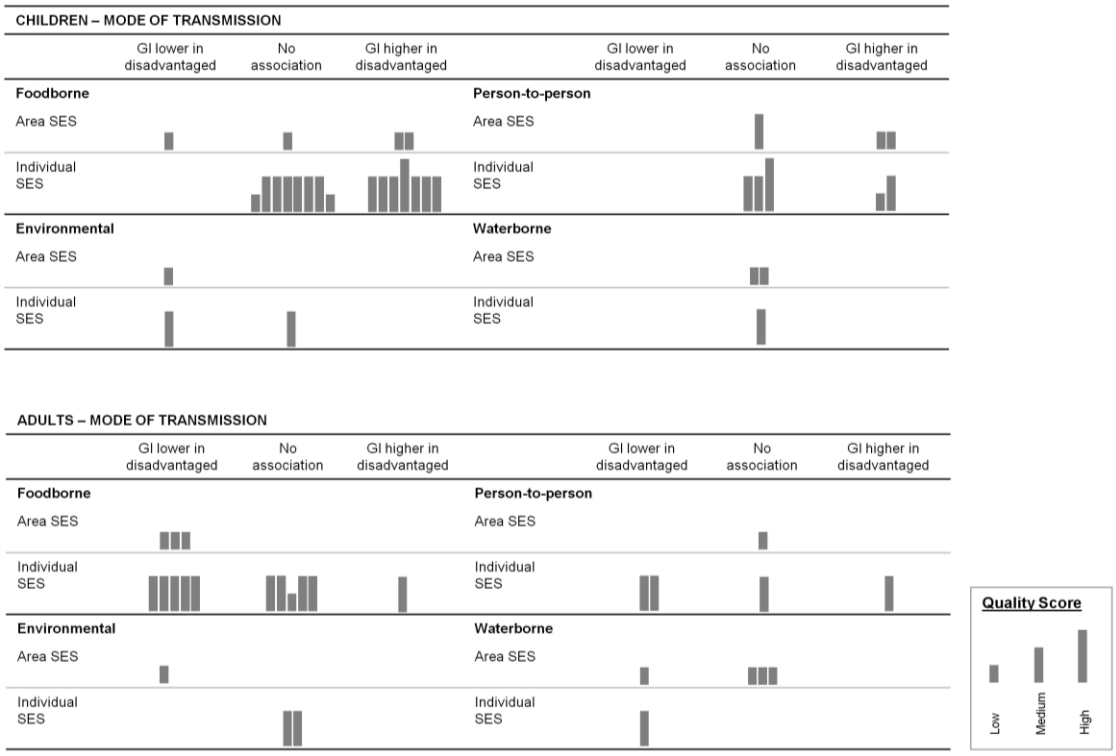


Legend
The harvest plots partition the point estimates for the association between SES and GI infection risk into three categories: estimates that indicate GI infection risk is lower for those of lower SES; estimates that indicate GI infection risk is higher for those of lower SES; and estimates that indicate there is no relationship. Each bar represents one point estimate. The height of the bar reflects the quality score assigned to the study from which the point estimate originated. The point estimates were grouped according to the age of the participants under investigation (children <18 years or adults ≥18 years), the method used to sample cases (population-based surveys, laboratory records, hospital admissions or GP presentations) and the SES measure used (area-level or individual-level).

Amongst the point estimates based on cases with a laboratory report (displayed in Figure 4.3), there was no clear modifying role of pathogen type (based on the predominant mode of transmission) on the relationship between SES and GI infection risk, although there were some differences by age. Again, the point estimates for children are shown on the upper half of the plot and the estimates for adults are shown on the lower half. For children, the majority of point estimates for foodborne pathogens (*Campylobacter*, *Salmonella*, *Yersinia enterocolitica*) showed a higher risk of GI infection among disadvantaged children or no association, whereas for adults most point estimates for foodborne pathogens showed a lower risk of GI infection among disadvantaged adults or no association. For environmental (STEC) and waterborne (*Cryptosporidium*, *Giardia*) pathogens, there were no point estimates suggesting a higher risk of GI infection among disadvantaged adults or children, however this was based on small numbers.

No clear difference was observed in the relationship between SES and GI infection risk, when comparing point estimates based on area-level and individual-level SES measures, or when comparing point estimates from studies conducted in different countries (based on level of development or climate) (data not shown).

Figure 4.3 Harvest plot for risk of GI infection by SES, stratified by predominant pathogen transmission route, SES measure and age



Legend
See legend for Figure 4.2. The point estimates were grouped according to the age of the participants under investigation (children <18 years or adults ≥18 years), the predominant mode of transmission of the pathogen under investigation (person-to-person, foodborne, environmental or waterborne) and the SES measure used (area-level or individual-level).

Meta-analysis and meta-regression

A total of 77 studies were included in the meta-analysis contributing a total of 83 point estimates. Of the 25 studies that could not be included in a meta-analysis, 15 did not provide sufficient quantitative data, six did not use a dichotomous outcome (four used linear regression, two used Pearson/Spearman correlation) and four studies analysed the same cases as other studies (details found in Appendix 4). Since age was highlighted as a key potential effect modifier in the harvest plots, point estimates from the same study stratified by age were retained individually in the meta-analysis to allow investigation of this variable.

The pooled risk ratio for GI infection comparing low verses high SES for all studies combined was 1.06 (95% CI 0.95–1.19), with considerable statistical heterogeneity (I^2 99.08%). Random-effects meta-regression was performed in an attempt to quantitatively explain some of the heterogeneity; exploring the impact of potential effect modifiers (identified a priori) on the relationship between SES and GI infection risk. The univariate and multivariate results are presented in Table 4.3.

In univariate analysis, the risk of GI infection between low and high SES groups was on average not statistically significantly different between studies conducted in countries with different climates and levels of development (Table 4.3). The risk of GI infection for low versus high SES was non-significantly lower among studies that used area-level compared to individual-level SES measures. Additionally, the risk of GI infection for low compared to high SES was non-significantly higher among studies that analysed cases identified via population-based surveys and general practices, and significantly higher among studies that analysed hospitalised cases, compared to studies that analysed laboratory recorded cases.

Amongst the studies that analysed cases with a laboratory record, the risk of GI infection for low compared to high SES was significantly lower among studies that analysed environmental pathogens, and significantly higher among studies that analysed pathogens transmitted via the person-to-person route, compared to studies that analysed foodborne pathogens.

In multivariate analysis (excluding pathogen type since not all studies analysed specific pathogens), age was identified as the only statistically significant potential effect modifier. The risk ratios for GI infection between low and high SES groups observed by studies that analysed children, were on average 1.87 times the risk ratios observed by studies that focused on adults, controlling for the other study differences included in the model (Table 4.3).

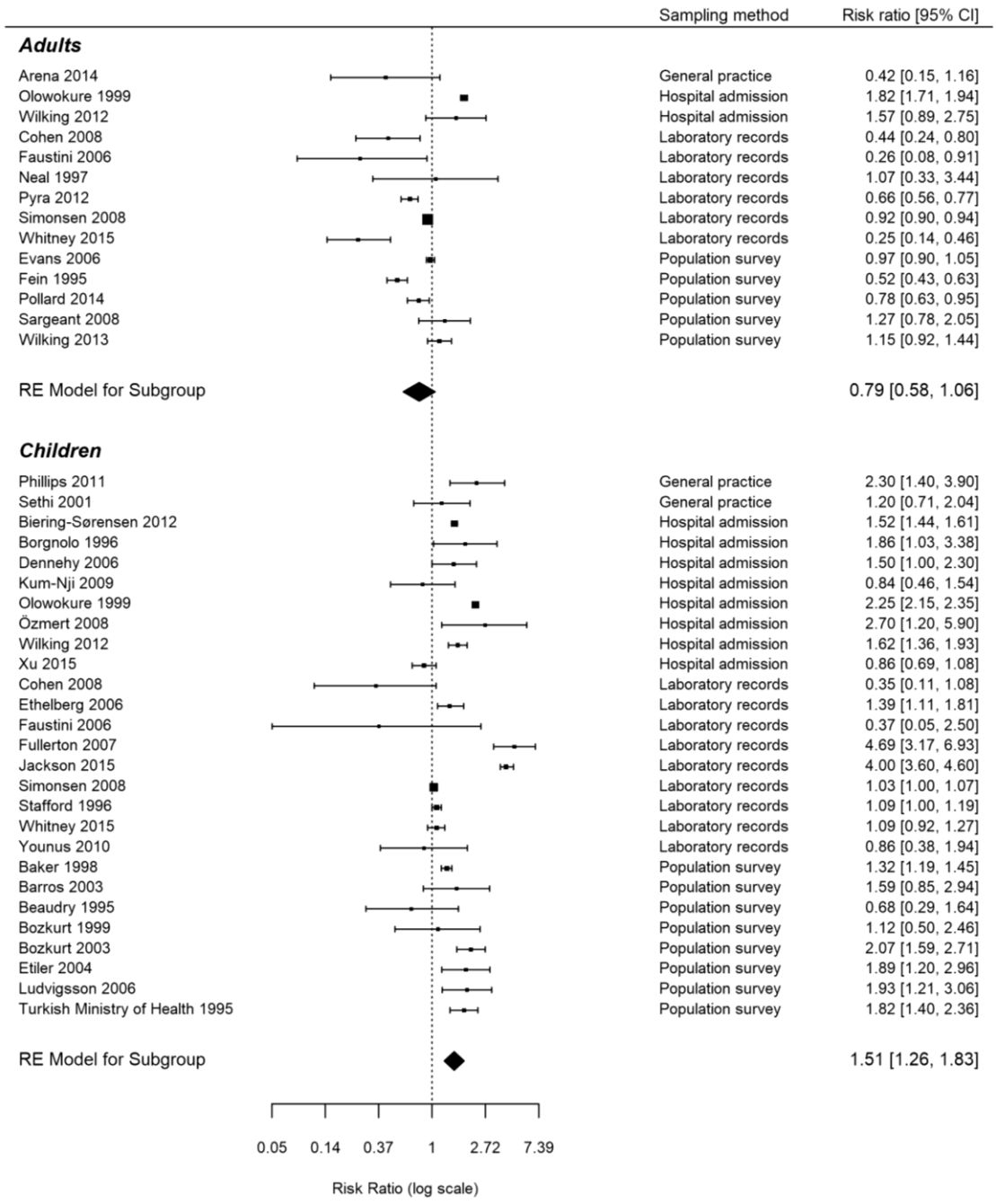
Table 4.3 Univariate and multivariate meta-regression results for GI infection risk between low and high SES groups

		Univariate RRR (95% CI)	Multivariate RRR (95% CI)	Number observations
Method of sampling GI infection cases	Laboratory records	1 (ref)	1 (ref)	43
	Population-based survey	1.11 (0.85–1.44)	1.04 (0.75–1.43)	23
	GP presentation	1.18 (0.71–1.94)	1.02 (0.62–1.69)	5
	Hospital admissions	1.49 (1.08–2.07)*	1.24 (0.88–1.73)	12
SES measure	Individual level	1 (ref)	1 (ref)	50
	Area level	0.87 (0.69–1.09)	0.92 (0.70–1.22)	33
Age of participants	Adult	1 (ref)	1 (ref)	14
	Mixed ages	1.17 (0.88–1.54)	1.22 (0.90–1.66)	42
	Child	1.89 (1.40–2.55)***	1.87 (1.35–2.59)***	27
Country Human Development Index[a]	Upper tertile	1 (ref)	1 (ref)	39
	Middle tertile	0.98 (0.76–1.25)	1.09 (0.84–1.41)	30
	Lower tertile	1.04 (0.73–1.49)	0.88 (0.62–1.25)	14
Country climate	Temperate/Mediterranean	1 (ref)	1 (ref)	62
	Arid	1.05 (0.69–1.61)	1.01 (0.67–1.52)	7
	Snow	0.81 (0.60–1.10)	0.89 (0.67–1.19)	14
Pathogen type[b]	Foodborne	1 (ref)	-	28
	Waterborne	0.73 (0.46–1.14)	-	8
	Environmental	0.46 (0.23–0.91)*	-	3
	Person-to-person	1.65 (1.05–2.59)*	-	7

*p <0.05; **p <0.01; ***p <0.001
[a] Higher values indicate higher level of human development
[b] Not all studies analysed specific pathogens, therefore this variable was not entered into the multivariate model
CI = confidence interval; GI = gastrointestinal; ref = reference category; RRR = ratio of risk ratios; SES = socioeconomic status

A forest plot for the studies stratified by age is shown in Figure 4.4. For children, the pooled risk ratio was 1.51 (95% CI 1.26–1.83) with I^2 97.87%. For adults, the pooled risk ratio was 0.79 (95% CI 0.58–1.06) with I^2 98.64%.

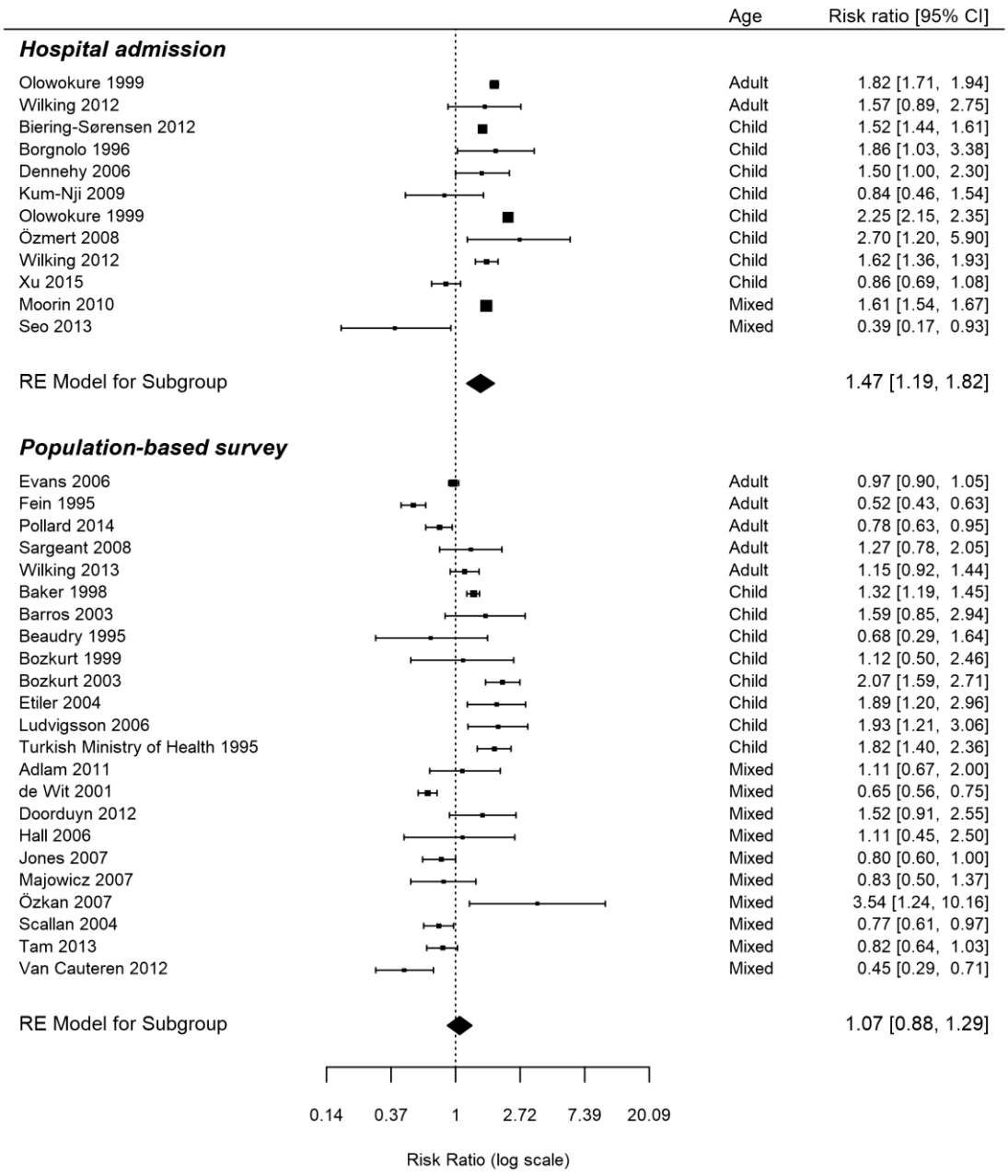
Figure 4.4 Forest plot for studies measuring GI infection risk between low and high SES groups, stratified by age



As mentioned, in multivariate meta-regression, age was identified as the only statistically significant potential effect modifier of the relationship between SES and GI infection risk. Yet, as highlighted in the harvest plot (Figure 4.2), the majority of point estimates calculated using cases identified via hospitals, were also based on children. To demonstrate this, a forest plot stratifying studies according to whether they sampled cases from hospitals or population-based surveys is shown in Figure 4.5. For studies that identified cases via

hospitals the pooled risk ratio was 1.47 (95% CI 1.19–1.82) with I^2 98.03%, and for population-based surveys the pooled risk ratio was 1.07 (95% CI 0.88–1.29) with I^2 92.54%. However, as can be seen, 67% of the studies that analysed hospitalised cases also analysed children (8 out of 12 studies), compared to 35% of the studies that analysed cases identified via population-based surveys (8 out of 23 studies).

Figure 4.5 Forest plot for studies measuring GI infection risk between low and high SES groups, stratified by method of sampling cases (population-based surveys or hospital admissions)



Sensitivity analysis

A number of sensitivity analyses were performed to check the robustness of the results. Very similar results were observed compared to those in the main analysis.

The meta-analyses for the child and adult subgroups were repeated using studies given a high and medium quality rating only. For adults, this resulted in a pooled RR of 0.82 (95% CI 0.64–1.05) from eight estimates and the I^2 was 95.78%. For children, this resulted in a pooled RR of 1.47 (95% CI 1.03–2.09) from nine estimates and the I^2 was 98.26%.

The meta-analyses for the child and adult subgroups were repeated excluding the studies where we calculated a point estimate from the raw data reported in the study. For adults, this resulted in a pooled RR of 0.79 (95% CI 0.63–0.97) from eight estimates and the I^2 was 94.26%. For children, this resulted in a pooled RR of 1.54 (95% CI 1.22–1.95) from 16 estimates and the I^2 was 97.05%.

The meta-analyses for the child and adult subgroups were repeated excluding population-based studies that reported an odds ratio. For adults, this resulted in a pooled RR of 0.70 (95% CI 0.43–1.14) from nine estimates and the I^2 was 99.11%. For children, this resulted in a pooled RR of 1.49 (95% CI 1.19–1.88) from 22 estimates and the I^2 was 98.50%.

The meta-analyses for the child and adult subgroups were repeated using only studies that statistically adjusted for potential confounding variables. For adults, this resulted in a pooled RR of 0.82 (95% CI 0.66–1.03) from nine estimates and the I^2 was 94.74%. For children, this resulted in a pooled RR of 1.47 (95% CI 1.12–1.94) from 13 estimates and the I^2 was 98.27%.

The meta-analyses for the child and adult subgroups were repeated without the studies that reported point estimates that were adjusted for other SES measures either statistically or by matching in case-control studies. For adults, this resulted in a pooled RR of 0.71 (95% CI 0.43–1.19) from eight estimates and the I^2 was 97.27%. For children, this resulted in a pooled RR of 1.45 (95% CI 1.16–1.82) from 18 estimates and the I^2 was 97.54%.

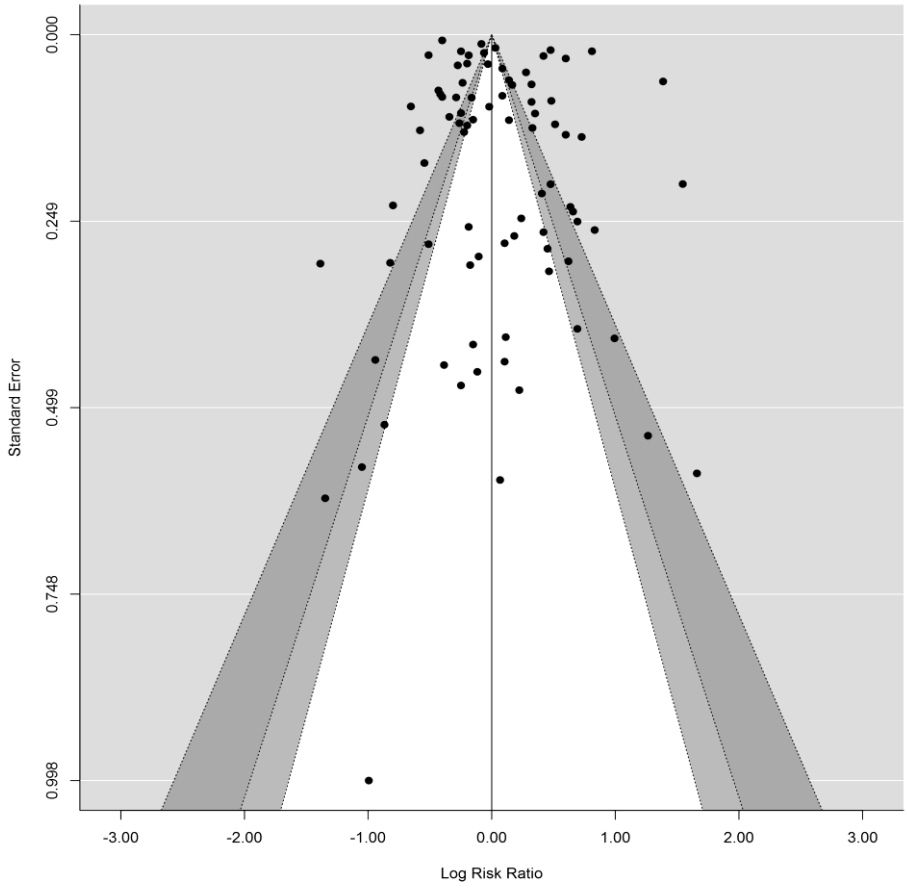
In the child subgroup three studies provided estimates for different pathogens. These studies used the same individuals for the denominators when calculating risk ratios for each pathogen, and therefore these non-cases were counted more than once in the meta-analysis. To assess the impact of double counting non-cases, sensitivity analysis was performed

whereby only one estimate from each multi-pathogen study was entered into the meta-analysis. The estimate for the pathogen with the largest number of cases was chosen within each study (i.e. *Giardia* for Cohen et al., 2008; *Campylobacter* for Simonsen, Frisch and Ethelberg, 2008; *Salmonella* for Whitney et al., 2015). The results for the child subgroup were very similar to those observed in the main analysis. The pooled RR was 1.54 (95% CI 1.28–1.84) and the I^2 was 97.26%.

Publication bias

A contour-enhanced funnel plot was produced to assess publication bias (Figure 4.6). The points within the plot appeared to be largely symmetrical, indicating that publication bias was unlikely.

Figure 4.6 Contour-enhanced funnel plot for included studies



White region = p-value >0.10; Mid-gray region = p-value between 0.10 and 0.05; Dark-gray region = p-value between 0.05 and 0.01; Region outside of funnel = p-value <0.01

4.6 DISCUSSION

In this systematic review and meta-analysis of observational studies from high income countries, evidence of an association between lower SES and a higher risk of GI infections for children, but not for adults, was observed. The pooled results indicate that disadvantaged children are 1.51 (95% CI 1.26–1.83) times as likely to experience a GI infection compared to their more advantaged counterparts. In adults, the pooled RR was 0.79 (95% CI 0.58–1.06). However, age explained a very small proportion of the heterogeneity observed across the studies as a whole.

The findings may reflect differential exposure, susceptibility or immunity to GI infections by SES in children. Children of lower SES might be more likely to come into contact with GI pathogens, or they may be more vulnerable to infection, possibly due to socially patterned factors that impair immune response, such as lack of breast feeding or nutritional deficiencies (Oakley et al., 2013; Lund and O'Brien, 2011; Darmon and Drewnowski, 2008; Jackson and Nazar, 2006). Antibodies that are produced in response to an infection, can provide protection against future disease for several years following the initial infection (Simmons et al., 2013; Miller et al., 2005). Thus, children of lower SES might be more exposed or susceptible to GI infections, but re-exposure may lead to better immunity and subsequent asymptomatic disease later in life.

To the best of my knowledge, only one other systematic review has been conducted on a similar topic previously. However, this previous review did not investigate results by age group. Newman et al. (2015) conducted a systematic review and narrative synthesis to investigate the association between SES and the risk of laboratory confirmed foodborne GI infections in high income countries. The authors identified 16 studies, and concluded that the association between SES and foodborne GI infection risk differed by pathogen. For most of the studies that examined *Campylobacter* and *Salmonella*, low SES compared to high was associated with a reduced risk of infection, however for *Listeria* low SES was associated with an increased risk. This finding was however based on one identified study examining cases of listeriosis, which highlights the limitations of investigating potential modifying factors using a small number of studies. Furthermore, a key limitation of the Newman et al. review was the narrow focus on laboratory-based studies. Cases of GI infection identified via laboratories represent a small fraction of cases occurring in the community (Tam et al., 2012b) and these cases may differ in terms of SES, to community cases.

The systematic review presented in this thesis was more broadly focused, and included studies that identified GI infection cases via population-based surveys, GPs and hospitals, as well as laboratory records. In this way, the review had greater scope for capturing the true burden of GI infections in the community. Additionally, it was possible to investigate the effects of several potential modifying factors of the association between SES and GI infection risk, such as the age of the participants under study, the country the study was conducted in, the measure of SES used and the methods used to sample the cases. These factors were specified a priori, based on knowledge of the subject matter, thus minimising the potential of drawing false positive conclusions due to 'data dredging' or performing multiple analyses of the data (Thompson and Higgins, 2002). The use of harvest plots as well as meta-analysis allowed all of the studies to be captured in the data synthesis, not exclusively studies with a quantitative measure. Selection bias was mitigated by double screening throughout, and the potential for publication bias was not evident in the funnel plot. Additionally, the included studies were conducted in a number of different high income countries which may have improved the external validity of the findings.

Nonetheless, there were several limitations to the analysis. Non-English language studies were excluded due to time limitations and costs of translating studies, and countries that have been in transition between middle and high income (e.g. Turkey) were included, both of which could potentially limit the interpretation of the results. The majority of the studies identified were assessed as being of generally low quality with potential for bias, which may have also biased the results of the review. For example, several case-control studies selected controls based on the geographical residence of cases or through case-nomination, thereby potentially biasing the relationship between SES and GI infection risk towards the null. However, meta-analysis results for the age subgroups were similar when sensitivity analyses were conducted excluding studies which controlled for or matched by SES, and when low quality studies were excluded.

Additionally, for several studies it was necessary to calculate a point estimate from the raw data provided within the study, and for studies that analysed SES as a continuous variable, a plausible low versus high SES comparison was estimated. To ensure that these calculations had not biased or influenced the main results of the review, a sensitivity analysis was conducted which excluded studies where estimates were calculated from the raw data. The results of the sensitivity analysis confirmed those of the review.

Odds ratios and relative risks for the association between SES and GI infection risk were combined in the meta-analysis, however odds ratios when interpreted as relative risks can

overstate the effect size (Davies, Crombie and Tavakoli, 1998). This can be especially problematic when disease incidence is ≥20%, and therefore odds ratios derived from population-based surveys may, in particular, have overstated the effect size. However, in sensitivity analysis excluding population-based surveys reporting odds ratios, pooled results for the child subgroup were similar, and for adults the pooled result (RR 0.70; 95% CI 0.43–1.14) was in fact smaller than the result observed in the main analysis, suggesting that the inclusion of population-based surveys reporting odds ratios had not lead to an overstatement of effect sizes in the main analysis.

Despite investigating several potential effect modifiers of the association between SES and GI infection risk, a considerable amount of heterogeneity in the effect estimates remained unexplained. It is possible that factors that could not be adjusted for may explain the high residual heterogeneity. The primary aims of the individual studies varied, as did the variables used to statistically adjust the associations between SES and GI infection risk. Additionally, the categorisation of low and high SES may have differed considerably between studies, for example, certain studies may have compared the bottom decile of an SES exposure versus the top decile, and others the bottom third versus the top third of the SES exposure distribution. All of these factors may have contributed to the high residual heterogeneity observed.

Unfortunately, it was not possible to include pathogen type in the multivariate meta-regression model because not all studies analysed specific pathogens. There was some evidence to suggest that the risk of GI infection for low compared to high SES was on average lower among studies that analysed pathogens commonly transmitted via the environment, and higher among studies that analysed pathogens transmitted person-to-person, compared to studies that analysed foodborne pathogens. However, these findings were based on small numbers; there were three and seven point estimates available for environmental and person-to-person pathogens, respectively. Individuals of lower SES are generally more likely to live in crowded housing (Solari and Mare, 2012), which could indicate that they are more likely to be exposed to pathogens commonly transmitted via the person-to-person route (such as norovirus, rotavirus and *Shigella*), compared to pathogens transmitted via alternative routes. Additionally, levels of deprivation are generally lower in rural areas (Gartner et al., 2008), and rurality has been cited as an important risk factor for STEC infection (Byrne et al., 2015). This might suggest that individuals of lower SES are less likely to be exposed to environmental pathogens such as STEC. These theories are however anecdotal and further research is warranted.

Finally, a key objective of this review was to explore potential effect modifiers of the relationship between SES and GI infection risk, in particular, to see whether the method of sampling cases (i.e. via population-based surveys, laboratory records, GP presentations or hospital admissions) modified the relationship. For instance, if similar social gradients in GI infection risk were reported by population-based surveys and studies that analysed hospitalised cases, this may have lent support to the hypothesis that individuals of lower SES have a greater need for healthcare services because they have a greater baseline risk of infection. When the results of studies that identified cases via hospitals were pooled, a statistically significant increased risk of GI infection for those of lower SES was observed (RR 1.47; 95% CI 1.19–1.82), whereas a non-significant increased risk was found when the results of population-based surveys were pooled (RR 1.07; 95% CI 0.88–1.29). However, the majority of studies that analysed hospitalised cases also analysed children.

The meta-regression results suggested no statistically significant modifying effect of the methods used to sample cases, when the age of the participants was accounted for. It is important to note, however, that the large amount of statistical heterogeneity observed may have negatively affected the power to detect statistically significant modifiers in the meta-regression (Hempel et al., 2013), and therefore non-significance should not necessarily be interpreted as evidence that a potential modifier had no effect on the relationship between SES and GI infection risk. Only two studies analysed hospital admissions in adults (Wilking et al., 2012; Olowokure et al., 1999), and both found an increased risk of hospital admission amongst those of lower SES, although the estimate for Wilking et al. (2012) did not reach statistical significance. These studies suggest that factors in addition to age may contribute towards an increased risk of GI infection requiring hospitalisation for those of lower SES.

Further research is needed to better understand inequalities in the risk of GI infection requiring hospitalization, especially in relation to whether the risk differs between adults and children.

Chapter 5

Results: Study 2

Relationship between SES and symptom severity and sickness absence in people with infectious intestinal disease in the UK

5.1 ABSTRACT

Background

Several studies conducted in high income countries suggest individuals of low SES compared to high have higher rates of GP consultation and hospital admission due to IID. The mechanisms explaining these apparent health inequalities are not completely understood, however contributing factors may include differential risk of infection, differential healthcare-seeking behaviour, or differential disease severity across socioeconomic groups. This study sought to investigate one of these potential mechanisms by examining the extent of socioeconomic inequalities in measures of IID severity.

Methods

A cross-sectional analysis was performed on individuals with IID identified from a large population-based survey (the UK-based IID2 study), to explore the associations between SES and symptom severity, SES and sickness absence, and to assess the role of symptom severity on the relationship between SES and absence. Regression modeling was used to examine these associations, whilst investigating the effects of several covariates such as age, sex, ethnicity, urban/rural residency and recent foreign travel.

Results

Among 1164 IID cases aged ≥5 years, those of lower versus high SES had twice the odds of experiencing severe symptoms (OR 2.2; 95% CI 1.66–2.87). Lower SES was associated with higher odds of sickness absence (OR 1.8; 95% CI 1.26–2.69), however this association was attenuated after adjusting for symptom severity (OR 1.4; 95% CI 0.92–2.07).

Conclusions

In a large sample of individuals with IID, those of low versus high SES were more likely to report severe symptoms, and sickness absence; with greater symptom severity largely explaining the higher absence. Further research is required to understand the mechanisms explaining greater severity of illness in disadvantaged groups, and to identify ways to minimise the differential impact of IID on sickness absence.

5.2 INTRODUCTION

Some studies conducted in high income countries, have found that individuals of low SES compared to high have higher rates of GP consultation (Phillips et al., 2011; Beale et al., 2010; Teschke et al., 2010; Quigley et al., 2006) and hospital admission due to IID (Biering-Sørensen et al., 2012; Lal et al., 2012; Wilking et al., 2012; Pockett et al., 2011; Moorin et al., 2010; Ma, El Khoury and Itzler, 2009; Özmert, Kilic and Yurdakök, 2008; Dennehy et al., 2006; Olowokure et al., 1999; Borgnolo et al., 1996). For example, in the West Midlands in the UK, hospital admission rates for young children with IID were twice as high in the most deprived areas compared to the least (Olowokure et al., 1999). However, the mechanisms explaining these apparent health inequalities are not well understood. Contributing factors may include differential risk of infection, differential healthcare-seeking behaviour, or differential disease severity across socioeconomic groups. Separating out the effects of these potential explanations is imperative to understand the role they play in generating the inequalities observed, and so that interventions and policies can be developed to tackle the problem.

This study examines the extent of socioeconomic inequalities in measures of IID severity. A previous cross-sectional analysis of IID cases aged ≥16 years identified in the English IID1 study, showed that IID cases of lower SES (as measured by educational attainment) were more likely to present to their GP for an episode of IID, compared to those of higher SES (Tam, Rodrigues and O'Brien, 2003). In addition, disease severity was strongly predictive of GP presentation for IID, however numbers were insufficient to assess the relationship between SES and IID severity. These findings indicate that healthcare-seeking behaviour for IID may be socially patterned, which potentially could be related to disease severity. Studies conducted in the Netherlands, France, New Zealand and the USA have also found disease severity to be a predictor of primary care presentation for IID for individuals of all ages (Doorduyn, Van Pelt and Havelaar, 2012; Van Cauteren et al., 2012; Adlam et al., 2011; Scallan et al., 2006; De Wit et al., 2001b).

Sickness absence might also be thought of as a measure of disease severity. Rates of general (all cause) sickness absence, have been shown to be higher for those of lower SES compared to high (Kristensen et al., 2010), however some studies have demonstrated that this association can in part be explained by the increased levels of morbidity for those of lower SES (Kaikkonen et al., 2015; Johansson and Lundberg, 2009). The few studies that have investigated the relationship between SES and sickness absence due to IID have produced

inconsistent results (Mohren et al., 2005; Feeney et al., 1998); and I am yet to find a study that has examined the role of IID symptom severity on the relationship between SES and sickness absence.

To gain a better understanding of inequalities in disease severity as a consequence of IID, I analysed a large sample of individuals with IID aged ≥5 years extracted from the UK-based IID2 study, to investigate the association between SES and measures of self-reported IID symptom severity and sickness absence.

5.3 AIM AND OBJECTIVES

Aim

- To investigate the association between SES and measures of self-reported IID severity using data collected in the population-based IID2 study in the UK.

Objectives

- To investigate the relationships between SES and self-reported:
  - IID symptom severity
  - Absence from work, school or daily activities due to IID
- To investigate the role of IID symptom severity on the relationship between SES and sickness absence

5.4 METHODS

The methods are described in detail in Chapter 3.

Ethical considerations

This analysis was performed using anonymised datasets. The original IID2 study was granted ethical approval by the North West Research Ethics Committee (07/MRE08/5) on 19th April 2007 (Tam et al., 2012b). Written informed consent was obtained from all participants and also the parent or guardian of child participants. When entering the study, participants gave consent for their anonymised data to be used for future analyses.

5.5 RESULTS: SAMPLE

Cohort study

The prospective Cohort study that was conducted as part of the IID2 study included 6836 participants. During follow-up participants reported 2276 episodes of diarrhoea and/or vomiting. Of these, 1409 participants completed a symptom questionnaire (62%). A detailed description of the characteristics of the participants who reported symptoms but did not complete a symptom questionnaire can be found in the final report of the IID2 study (Tam et al., 2012b). Briefly, among those who reported symptoms, males and those of Non-White ethnicity (compared to White ethnicity) were less likely to submit a questionnaire. Those aged from 5–44 years were less likely to submit a questionnaire compared to those aged over 65 years, however those aged from 0–4 (parent-report) and from 45–64 years were more likely. There was no evidence of a linear trend in the odds of questionnaire completion across IMD quintiles. Participants who reported symptoms in the most deprived quintiles were no more or less likely to submit a questionnaire compared to those in the least deprived quintile (Tam et al., 2012b).

The 1409 episodes of IID were experienced by 1145 individual participants. All 1145 cases had experienced either diarrhoea or vomiting. Characteristics of the Cohort study cases are shown in Table 5.1.

GP Presentation study

In total, 2203 patients were invited to participate in the GP Presentation study. Of these, 1264 (57%) attended a baseline interview, and 1254 (57%) were recruited. Participation was lowest amongst those aged 15–24 years and highest amongst those aged 55–64 years (Tam et al., 2012b). The vast majority of recruited participants completed a questionnaire. There were only 45 individuals (4% of recruited participants) who did not complete a questionnaire or the symptom information was absent or missing. In total, 77 individuals (6% of recruited participants) experienced their illness for 14 days or longer and were excluded (Tam et al., 2012b).

The 1132 episodes of IID were experienced by 1122 individual participants. All 1122 cases had experienced either diarrhoea or vomiting. Characteristics of the GP Presentation study cases are shown in Table 5.1.

Table 5.1 Characteristics of Cohort and GP Presentation study cases

Study	GP Presentation	Cohort
Number	1122	1145
Age in years, mean (SD)	38 (26.8)	44 (24.8)
Male	530 (47.2)	391 (34.1)
Ethnicity Non White	87 (7.8)	22 (1.9)
Residence		
Urban	828 (73.8)	791 (69.1)
Rural	291 (25.9)	354 (30.9)
NA	3 (0.3)	0
Foreign travel before illness		
Not travelled	981 (87.4)	1037 (90.6)
Travelled	140 (12.5)	96 (8.4)
NA	1 (0.1)	12 (1.0)
NS-SEC		
Managerial/professional	468 (41.7)	643 (56.2)
Intermediate	207 (18.4)	184 (16.1)
Routine/manual	268 (23.9)	148 (12.9)
NA	179 (16.0)	170 (14.8)
GP presentation for illness	1122 (100)	67 (5.9)

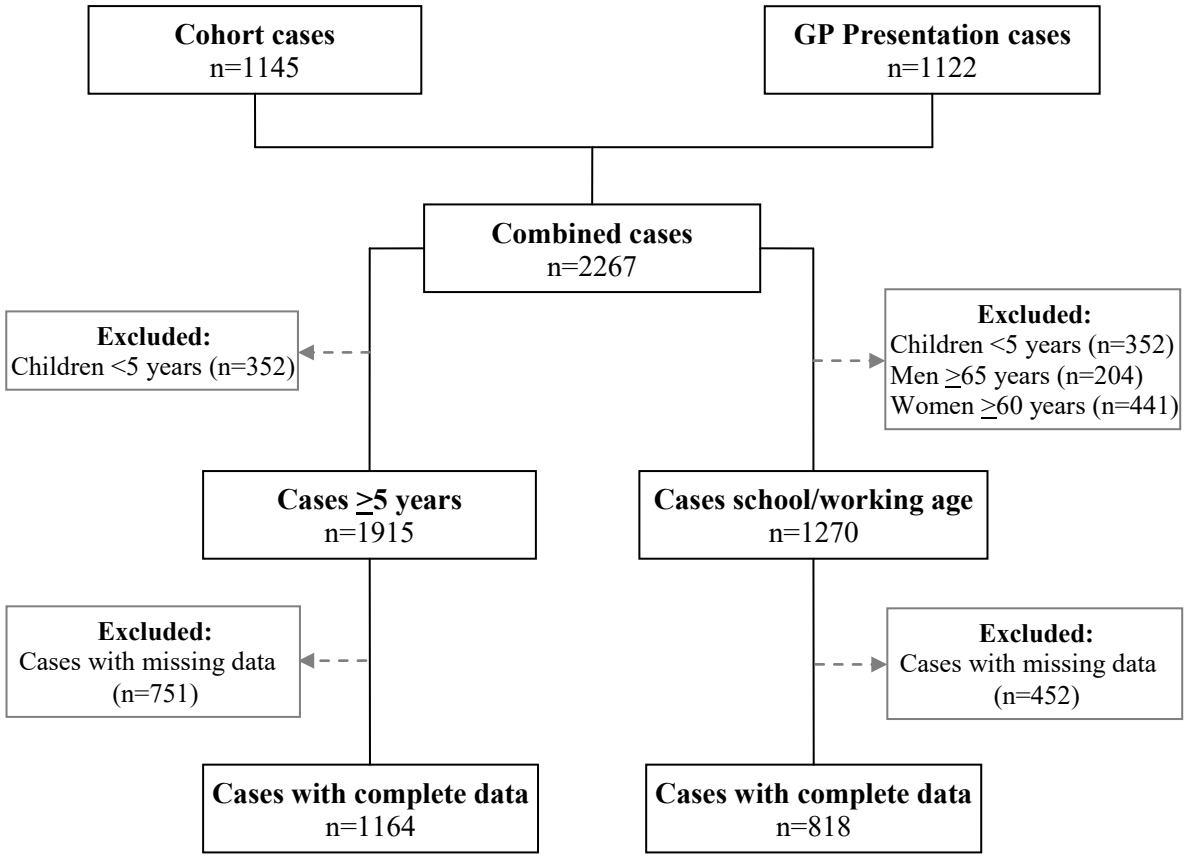
Figures expressed as number (%) except where stated otherwise
GP = general practice; NA = missing; NS-SEC = National Statistics Socioeconomic Classification; SD = standard deviation

As can be seen from Table 5.1, the characteristics of the cases from the two studies were largely comparable, with the exception of the proportion of cases who consulted their GP for their illness (100% and 5.9% for GP Presentation and Cohort studies, respectively). This large difference is a reflection of the design of the two studies, since all cases in the GP Presentation study were recruited following a consultation for IID.

IID cases from Cohort and GP Presentation studies combined

When the Cohort and GP Presentation study IID cases were combined there were 2267 cases available for analysis. For the symptom severity outcome, the analysis was performed using cases aged five years or over (n=1915). Cases of school or working age were used to investigate the sickness absence outcome (n=1270). A flowchart detailing the sampling process can be seen in Figure 5.1.

Figure 5.1 Flowchart of sampling process



5.6 RESULTS: IID SYMPTOM SEVERITY

Descriptive statistics

Characteristics of the IID cases aged five years or older are displayed in Table 5.2. Overall, the majority of the cases were in managerial/professional occupations (49.6%) and relatively few were in intermediate (17.2%) or routine/manual occupations (17.6%). The vast majority of the cases were of White ethnicity (96%), and over a third were male (37.9%). The majority of the cases resided in urban areas (71.1%), and had not travelled outside the UK in the ten days before the onset of their illness (87.5%).

Table 5.2 Characteristics of cases ≥5 years of age

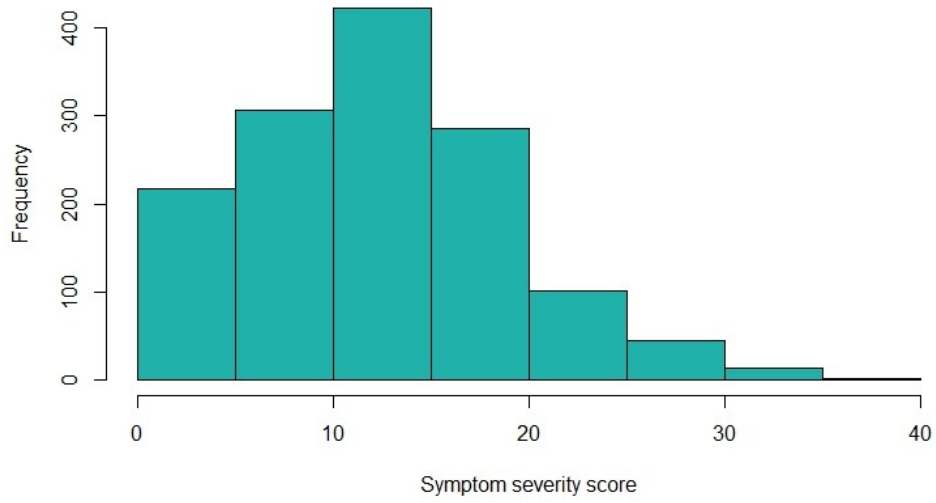
	Cases ≥5 years of age (n=1915)				
	Percentage within each category of NS-SEC				
	Managerial/ professional	Intermediate	Routine/ manual	p-value[a]	All cases[b]
	(n=949)	(n=330)	(n=337)		(n=1915)
Age group (years)					
5–14	12.2	9.4	9.5	0.342	10.6
15–24	4.4	5.2	7.4		5.2
25–44	24.1	22.7	23.7		22.9
45–64	36	36.7	33.5		35.1
65+	23.2	26.1	25.8		26.2
Male	38	32.7	45.1	0.004	37.9
Ethnicity Non-White	3.4	4.8	3.6	0.468	4
Rural residence	30.6	30	19	<0.001	28.8
Travelled before illness	14.3	10.3	7.7	0.004	11.9
Symptom severity					
Mild	38.3	34.4	20.3	<0.001	24.5
Moderate	34	33	35.5		24.9
Severe	27.7	32.6	44.2		23.4

IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; SD = standard deviation
Figures expressed as percentages except where stated otherwise
[a] Statistical significance of relationship between NS-SEC and each variable, tested using χ^2 test
[b] Total number of cases includes those with missing NS-SEC
Missing data (%): Urban/rural = 0.1; Foreign travel = 0.6; NS-SEC = 16; Symptom severity = 27

Cases in routine/manual compared to managerial/professional occupations were less likely to reside in rural areas (χ^2 17.4; p-value <0.001), be female (χ^2 10.9; p-value 0.004) or have travelled abroad before their illness (χ^2 11.2; p-value 0.004). Age and ethnicity were not statistically significantly associated with NS-SEC (Table 5.2).

The outcome variable, IID symptom severity score, could be calculated for 1395 cases (520 [27.2%] missing). The symptom severity score ranged from 2 to 40, with a median of 12 and mean of 12.8, and an interquartile range of 8 to 17. The positive skew of the severity score variable can be seen in Figure 5.2. The symptom severity score was converted into tertiles, the boundaries of which were: mild (severity score 2–9), moderate (severity score 10–15) and severe (severity score 16–40).

Figure 5.2 Histogram of IID symptom severity score for cases ≥ 5 years of age



Missing data

Of the exposure variables, urban/rural residency, travel and NS-SEC had missing data (0.1%, 0.6% and 15.6% respectively). As mentioned, for 27.2% of cases a symptom severity score could not be calculated. The nature of the missing data is explored in detail in Appendix 5, however for the NS-SEC and symptom severity variables, the characteristics of cases with missing data compared to those without missing data were largely similar. In total, 1164 (61%) cases had complete data for the variables of interest and were included in the univariate and multivariate analyses.

Univariate analysis

Figure 5.3 shows the relationship between the symptom severity score and the primary exposure of interest NS-SEC. As can be seen, the proportion of cases with severe symptoms increases as SES decreases from managerial/professional to routine/manual occupations.

Figure 5.3 Spineplot showing univariate relationship between IID symptom severity and NS-SEC

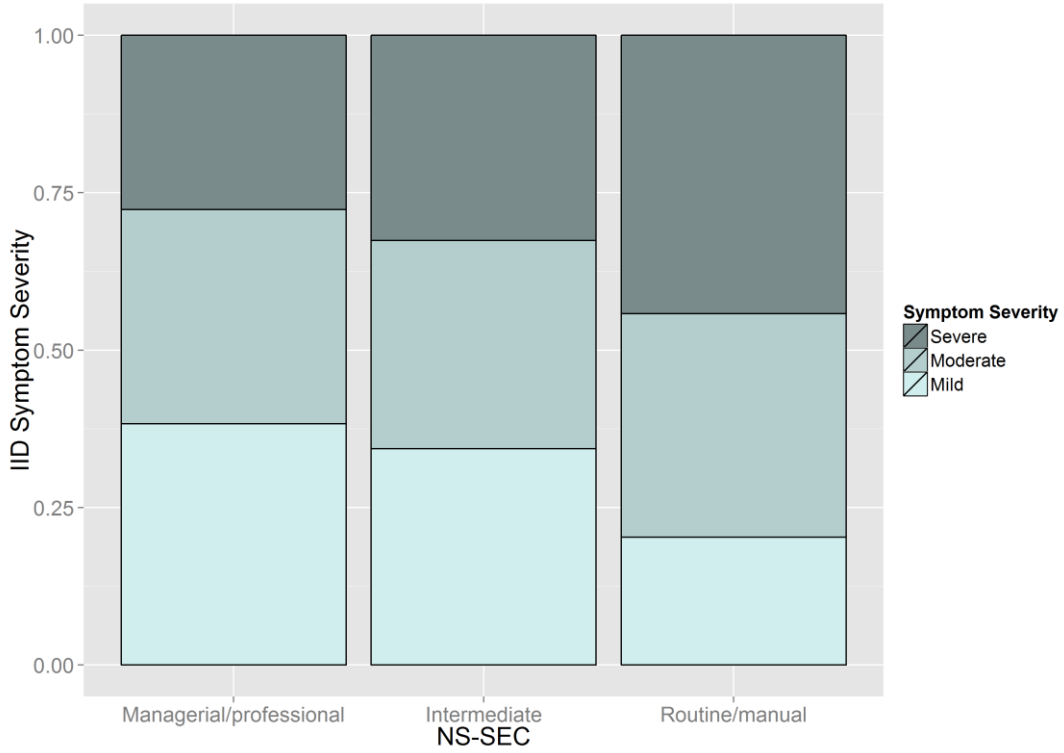


Table 5.3 shows the proportional odds ratios and 95% confidence intervals for the univariate relationships between the independent variables and the three category IID symptom severity dependent variable, using ordinal logistic regression. Age, ethnicity and NS-SEC were statistically significantly associated with symptom severity. Cases aged 15–24 years compared to 5–14 years, and those of Non-White ethnicity compared to White, had greater odds of experiencing severe IID symptoms versus mild or moderate symptoms combined, however these estimates were based on small numbers (61 cases were aged 15–24 years; 43 cases were of Non-White ethnicity). Cases aged 65 years and over, compared to 5–14 years had lower odds of experiencing severe symptoms, versus mild or moderate symptoms combined. Cases in routine/manual compared to managerial/professional occupations had

approximately two times the odds of experiencing severe symptoms versus mild or moderate symptoms combined.

Table 5.3 Univariate ordinal logistic regression for severe IID symptoms, versus mild or moderate symptoms combined for cases ≥5 years of age

	Severe symptoms versus mild or moderate symptoms combined OR (95% CI)
	Cases with complete data ≥5 years of age (n=1164)
Age group (years)	
5–14	reference
15–24	2.88 (1.59–5.31)
25–44	0.99 (0.68–1.45)
45–64	0.70 (0.49–1.01)
65+	0.60 (0.41–0.89)
Sex	
Female	reference
Male	0.92 (0.74–1.14)
Ethnicity	
White	reference
Non-White	2.27 (1.28–4.10)
NS-SEC	
Managerial/professional	reference
Intermediate	1.21 (0.92–1.61)
Routine/manual	2.18 (1.67–2.86)
Residence	
Urban	reference
Rural	0.82 (0.65–1.03)
Travelled before illness	
No	reference
Yes	1.21 (0.88–1.66)

CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

Multivariate analysis

Hierarchical ordinal logistic regression was performed to assess the relationship between SES and IID symptom severity. Table 5.4 shows the estimates and standard errors on the log odds scale for five models and their summary statistics. The dependent variable for all

models was IID symptom severity from mild to severe. Model 1 shows the results of the multivariate analysis with age, sex and ethnicity as the exposures. Model 2 shows the results with age, sex and ethnicity as the exposures, with the addition of NS-SEC as the primary exposure of interest. The addition of NS-SEC statistically significantly improved the model fit when comparing the likelihoods of Model 2 and 1 using the likelihood ratio chi-square statistic (Likelihood ratio χ^2 31.7; p-value <0.001) (Table 5.5).

Table 5.4 Nested multivariate ordinal logistic regression models for severe IID symptoms versus mild or moderate symptoms combined, for cases ≥5 years of age

	Model 1	Model 2	Model 3	Model 4	Model 5
Age 15–24 years [a]	1.10349***	0.99417**	0.99555**	0.96703**	0.96844**
std. error	0.3076	0.31034	0.31034	0.31079	0.31077
Age 25–44 years [a]	0.01956	-0.03688	-0.0398	-0.06495	-0.06746
std. error	0.19668	0.19769	0.19775	0.1984	0.19844
Age 45–64 years [a]	-0.31089	-0.37452*	-0.36862	-0.41046*	-0.40452*
std. error	0.18844	0.18957	0.18978	0.19083	0.19105
Age 65+ years [a]	-0.44237*	-0.50414*	-0.50386*	-0.51442*	-0.51405*
std. error	0.20123	0.20235	0.20241	0.20245	0.2025
Male [b]	-0.04751	-0.10106	-0.10177	-0.10966	-0.11028
std. error	0.11212	0.1134	0.11341	0.11355	0.11356
Ethnicity Non-White [c]	0.74436*	0.70799*	0.68846*	0.69326*	0.67497*
std. error	0.29843	0.30005	0.30129	0.29971	0.30091
NS-SEC Intermediate [d]		0.19235	0.19042	0.20174	0.19987
std. error		0.14423	0.14425	0.14436	0.14438
NS-SEC Routine/manual [d]		0.77800***	0.76855***	0.80001***	0.79092***
std. error		0.13947	0.14006	0.14026	0.14087
Rural residency [e]			-0.08963		-0.08452
std. error			0.12344		0.12356
Foreign travel [f]				0.27289	0.26995
std. error				0.16595	0.16599
Log-likelihood	-1255.1	-1239.3	-1239	-1237.9	-1237.7
Deviance	2510.2	2478.6	2478	2475.9	2475.4
AIC	2526.2	2498.6	2500	2497.9	2499.4
BIC	2566.7	2549.2	2555.7	2553.5	2560.1
Number	1164	1164	1164	1164	1164

*p <0.05; **p <0.01; ***p <0.001
[a] reference category = Age 5–14 years
[b] reference category = Female
[c] reference category = Ethnicity White
[d] reference category = NS-SEC Managerial/professional occupations
[e] reference category = Urban residency
[f] reference category = No foreign travel before illness
AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; NS-SEC = National Statistics Socioeconomic Classification; std. error = standard error

Table 5.5 Likelihood ratio tests for comparison of nested models

Test	Likelihood ratio χ^2 statistic	p-value
Model 1 versus Model 2	31.653	<0.001***
Model 2 versus Model 3	0.527	0.468
Model 2 versus Model 4	2.710	0.100
Model 2 versus Model 5	3.178	0.204

*p <0.05; **p <0.01; ***p <0.001

Table 5.6 Multivariate Model 2 for severe IID symptoms versus mild or moderate symptoms combined, for cases ≥5 years of age

	Severe symptoms versus mild or moderate symptoms combined OR (95% CI) Cases with complete data ≥5 years of age (n=1164)
Age group (years)	
5–14	reference
15–24	2.70 (1.48–5.02)
25–44	0.96 (0.65–1.42)
45–64	0.69 (0.47–1.00)
65+	0.60 (0.41–0.90)
Sex	
Female	reference
Male	0.90 (0.72–1.13)
Ethnicity	
White	reference
Non-White	2.03 (1.14–3.70)
NS-SEC	
Managerial/professional	reference
Intermediate	1.21 (0.91–1.61)
Routine/manual	2.18 (1.66–2.87)

CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

There was no improvement in the model fit when the variables urban/rural residency and recent foreign travel were added to Model 2 (Table 5.5). The proportional odds ratios and 95% confidence intervals for Model 2 are presented in Table 5.6. Those in routine/manual compared to managerial/professional occupations had approximately two times the odds of experiencing severe IID symptoms versus mild or moderate symptoms combined (OR 2.18; 95% CI 1.66–2.87) (Table 5.6). For those in intermediate compared to

managerial/professional occupations the odds of experiencing severe symptoms versus mild or moderate symptoms combined did not reach statistical significance (OR 1.21; 95% CI 0.91–1.61). Similar to the results from the univariate analysis cases aged 15–24 years compared to 5–14 years, and those of Non-White ethnicity compared to White, had greater odds of experiencing severe IID symptoms, and cases aged 65 years and over, compared to 5–14 years had lower odds of experiencing severe symptoms versus mild or moderate symptoms combined.

Assumptions

An assumption of ordinal logistic regression is that the coefficients describing the relationship between each pair of outcome categories are the same (the proportional odds assumption). A graphical method was used to test this assumption for Model 2 (displayed in Appendix 5). The plot showed that for various levels of the exposure variables, the difference between predicted logits for each category of the dependent variable were approximately similar, indicating the proportional odds assumption held true. This suggested that ordinal logistic regression was an appropriate method for modeling the symptom severity outcome.

Additionally, the appropriateness of combining cases from the IID2 component studies was supported by analyses indicating the relationship between NS-SEC and the symptom severity outcome was not statistically significantly different between the Cohort and GP Presentation studies (Appendix 5).

Sensitivity analyses

Similar results to those reported were observed when analyses were conducted with recurrent episodes of IID included with clustering at the individual level accounted for using mixed-effects models, and when linear regression was used with the log of the symptom severity score (Appendix 5). The results were also similar and in fact a stronger association between NS-SEC and symptom severity was observed when the boundaries of the three symptom severity categories were changed so that there was an equal 12 point severity score difference within each category, i.e. mild (severity score 2–14), moderate (severity score 15–27) and severe (severity score 28–40) (details in Appendix 5). There were however a small number of cases in the severe category using the altered boundaries (mild [n=875], moderate [n=486] and severe [n=34]).

Results from analyses involving cases of all ages (0 to 90+ years), and stratified results for children (aged ≥5 to <16 years) and adults (aged ≥16 years), also confirmed those from the main analyses (Appendix 5). The relationship between NS-SEC and symptom severity appeared to be strongest for children, although the number of child cases available to analyse was small (n=129). Children whose main household earner was in a routine/manual compared to managerial professional occupation had nearly three times the odds of experiencing severe IID symptoms versus mild or moderate symptoms combined (OR 2.96; 95% CI 1.22–7.48).

Results from the multiply imputed datasets (detailed in Appendix 5) confirmed those from this analysis using listwise deletion. The magnitude of the association between symptom severity and routine/manual compared to managerial/professional occupations was slightly weakened (OR 1.80; 95% CI 1.43–2.27) when the multiply imputed datasets were analysed. For those in intermediate compared to managerial/professional occupations the odds of experiencing severe IID symptoms versus mild or moderate symptoms combined did reach statistical significance when using the multiply imputed dataset with 1915 cases (OR 1.26; 95% CI 1.00–1.59). Ethnicity was not associated with symptom severity when analyses were performed using the imputed datasets.

5.7 Results: Sickness absence due to IID

Descriptive statistics

Characteristics of the IID cases of school or working age are shown in Table 5.7. Over half of the cases experienced absence from work, school or daily activities following an episode of IID (61%). Amongst the absentees, the majority were absent for 1–2 days (62%), and few were absent for more than five days (8%) (see Appendix 5 for analysis of absence duration).

Cases in routine/manual compared to managerial/professional occupations were less likely to reside in rural areas (χ^2 7.9; p-value 0.019), be female (χ^2 8.6; p-value 0.014) or have travelled abroad before their illness (χ^2 8.6; p-value 0.013). Age and ethnicity were not statistically significantly associated with NS-SEC (Table 5.7).

Table 5.7 Characteristics of cases of school or working age

	Cases school/working age (n=1270)				
	Percentage within each category of NS-SEC				
	Managerial/ professional (n=662)	Intermediate (n=215)	Routine/ manual (n=228)	p-value[a]	All cases[b] (n=1270)
Age (years) mean (SD)	37.3 (17.3)	38.7 (17)	37.7 (16.8)	0.586	37.7 (17.2)
Male	39.7	37.2	49.6	0.014	41
Ethnicity Non-White	3.9	7	4.8	0.185	5.1
Rural residence	30.3	28.4	20.6	0.019	27.6
Travelled before illness	15.6	13.1	7.9	0.013	12.9
Symptom severity					
Mild	40	36	20.7	<0.001	26.8
Moderate	36.7	32	38		27.1
Severe	23.3	32	41.3		22.1
Absent work/school	61.4	62.7	71.6	0.023	61

IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; SD = standard deviation
Figures expressed as percentages except where stated otherwise
[a] Statistical significance of relationship between NS-SEC and each variable, tested using χ^2 test and one-way analysis of variance (ANOVA) for age
[b] Total number of cases includes those with missing NS-SEC
Missing data (%): Urban/rural = 0.2; Foreign travel = 0.3; NS-SEC = 13; Symptom severity = 24; Absence = 2.7

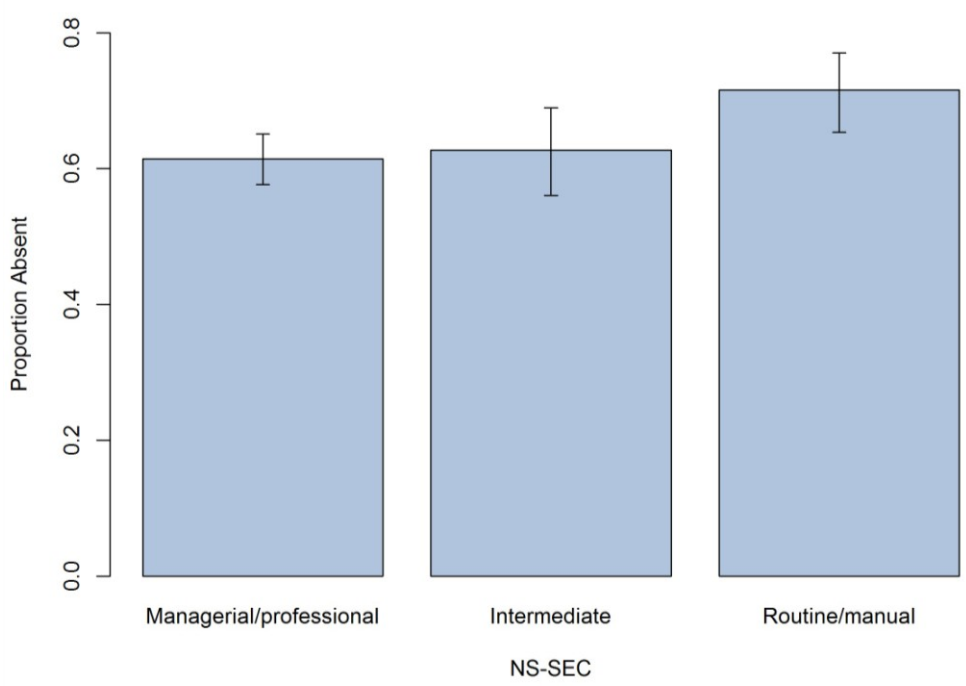
Missing data

The variables with missing data were absence, foreign travel, urban/rural residency, NS-SEC and symptom severity (2.7%, 0.3%, 0.2%, 13% and 24%, respectively). The nature of the missing data is explored in Appendix 5. The characteristics of cases with missing data compared to those without missing data were largely similar. A larger proportion of cases with missing sickness absence were of Non-White ethnicity compared to cases who answered the absence question (Appendix 5). Of the 1270 cases of school or working age, 818 (64%) had complete data for the variables of interest and were included in the univariate and multivariate analyses.

Univariate analysis

Figure 5.4 shows the relationship between the absence outcome and the primary exposure of interest NS-SEC. The proportion of those who were absent increased as SES decreased from managerial/professional to routine/manual occupations, and there was a statistically significant trend in the proportions (χ^2 6.54; p-value 0.01).

Figure 5.4 Proportion of absence for each NS-SEC group with error bars



Chi-squared Test for Trend in Proportions: $\chi^2 = 6.54$; p-value = 0.011

Table 5.8 shows the odds ratios and 95% confidence intervals for the univariate relationships between the independent variables and the binary sickness absence dependent variable, using logistic regression. Age, ethnicity, foreign travel, symptom severity and NS-SEC were statistically significantly associated with sickness absence due to IID. Those in routine/manual compared to managerial/professional occupations had 1.77 times the odds of absence (95% CI 1.22–2.58). Odds ratios only approximate the relative risk when the outcome is a rare event. The absence outcome was a common event and therefore the odds ratios do not approximate the relative risk in this analysis.

Table 5.8 Univariate logistic regression for sickness absence due to IID for cases of school/working age

	Sickness absence versus no sickness absence OR (95% CI)[a]
	Cases with complete data school/working age (n=818)
Age (years)	0.98 (0.97–0.99)
Sex	
Female	reference
Male	0.94 (0.70–1.25)
Ethnicity	
White	reference
Non-White	3.13 (1.38–8.41)
NS-SEC	
Managerial/professional	reference
Intermediate	1.13 (0.78–1.66)
Routine/manual	1.77 (1.22–2.58)
Residence	
Urban	reference
Rural	0.98 (0.71–1.34)
Travelled before illness	
No	reference
Yes	0.66 (0.44–0.99)
Symptom severity	
Mild	reference
Moderate	3.88 (2.75–5.51)
Severe	5.99 (4.07–8.95)

CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

[a] Since the absence outcome was common, the odds ratios should not be interpreted as relative risks

Multivariate analysis

Hierarchical multivariate logistic regression was performed to assess the relationship between SES and the binary sickness absence outcome. Table 5.9 reports the estimates and standard errors on the log odds scale for six nested models and their summary statistics for comparison. The dependent variable for all models was sickness absence from work, school or daily activities due to an episode of IID. Model 1 shows the results of multivariate logistic

regression with age, sex and ethnicity as the exposures. Model 2 shows the results with age, sex and ethnicity as the exposures, with the addition of NS-SEC as the primary exposure of interest. The addition of NS-SEC statistically significantly improved the model fit when comparing the likelihoods of Model 2 and 1 using the likelihood ratio chi-square statistic (Likelihood ratio χ^2 10.2; p-value 0.006) (Table 5.10). Those in routine/manual compared to managerial/professional occupations had greater odds of absence (OR 1.8; 95% CI 1.26–2.69) (Table 5.11). There was no improvement in the model fit when the variables urban/rural residency and recent foreign travel were added to Model 2 (Table 5.10).

Table 5.9 Nested multivariate logistic regression models for sickness absence due to IID for cases of school/working age

	Model1	Model2	Model3	Model4	Model5	Model6
Age (years)	-0.01560***	-0.01617***	-0.01653***	-0.01561***	-0.01594***	-0.01208*
std. error	0.00449	0.00451	0.00455	0.00454	0.00457	0.00473
Male [a]	-0.04642	-0.09753	-0.09816	-0.09156	-0.09226	-0.085
std. error	0.14892	0.15081	0.15086	0.15106	0.15111	0.15976
Ethnicity Non-White [b]	0.97973*	0.94604*	0.97140*	0.96563*	0.98776*	0.64454
std. error	0.45813	0.46022	0.46163	0.46089	0.46216	0.47473
NS-SEC Intermediate [c]		0.1196	0.11672	0.1151	0.11262	0.05006
std. error		0.19576	0.1959	0.196	0.19613	0.20843
NS-SEC Routine/manual [c]		0.60445**	0.61551**	0.57632**	0.58701**	0.31961
std. error		0.19388	0.19449	0.19482	0.19551	0.20674
Rural residency [d]			0.12445		0.11282	
std. error			0.16641		0.1667	
Foreign travel [e]				-0.30088	-0.2939	
std. error				0.20822	0.2086	
Symptoms Moderate [f]						1.28210***
std. error						0.17932
Symptoms Severe [f]						1.66116***
std. error						0.20556
Log-likelihood	-531	-525.9	-525.7	-524.9	-524.7	-482.4
Deviance	1062	1051.9	1051.3	1049.8	1049.3	964.9
AIC	1070	1063.9	1065.3	1063.8	1065.3	980.9
BIC	1088.9	1092.1	1098.3	1096.7	1103	1018.5
Number	818	818	818	818	818	818

*p <0.05; **p <0.01; ***p <0.001
[a] reference category = Female
[b] reference category = Ethnicity White
[c] reference category = NS-SEC Managerial/professional occupations
[d] reference category = Urban residency
[e] reference category = No foreign travel
[f] reference category = Symptom severity Mild
AIC = Akaike information criterion; BIC = Bayesian information criterion; NS-SEC = National Statistics Socioeconomic Classification; std. error = standard error

Table 5.10 Likelihood ratio tests for comparison of nested models

Test	Likelihood ratio χ^2 statistic	p-value
Model 1 versus Model 2	10.164	0.006**
Model 2 versus Model 3	0.562	0.453
Model 2 versus Model 4	2.072	0.150
Model 2 versus Model 5	2.532	0.282
Model 2 versus Model 6	87.01	<0.001***

*p <0.05; **p <0.01; ***p <0.001

Table 5.11 Multivariate Models 2 and 6 for sickness absence due to IID for cases of school/working age

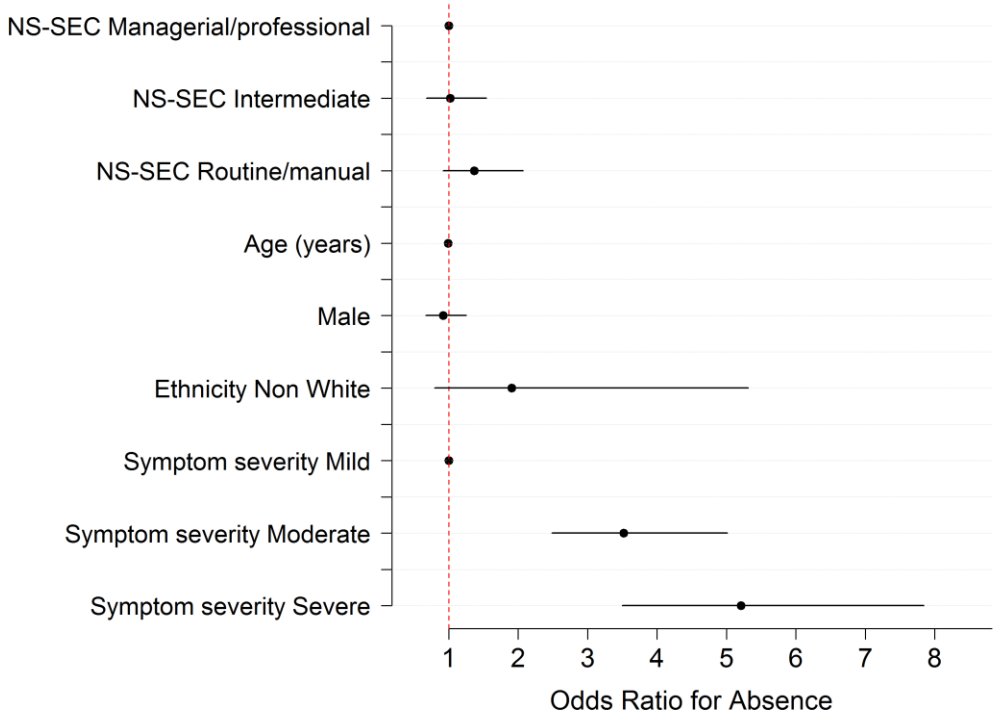
	Model 2	Model 6
	OR (95% CI)[a]	OR (95% CI)[a]
Age (years)	0.98 (0.98–0.99)	0.99 (0.98–1.00)
Sex		
Female	reference	reference
Male	0.91 (0.68–1.22)	0.92 (0.67–1.26)
Ethnicity		
White	reference	reference
Non-White	2.58 (1.12–7.00)	1.91 (0.80–5.31)
NS-SEC		
Managerial/professional	reference	reference
Intermediate	1.13 (0.77–1.66)	1.05 (0.70–1.59)
Routine/manual	1.83 (1.26–2.69)	1.38 (0.92–2.07)
Symptom severity		
Mild		reference
Moderate		3.60 (2.54–5.14)
Severe		5.27 (3.54–7.93)

CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio
[a] Since the absence outcome was common, the odds ratios should not be interpreted as relative risks
Cases with complete data school/working age (n=818)

A final model was created to investigate the effect of IID symptom severity on the absence outcome. The odds ratios and 95% confidence intervals for Models 2 and 6 are presented in Table 5.11. When symptom severity was added to Model 2 the odds of absence for those in routine/manual compared to managerial/professional occupations was attenuated and rendered non-significant (OR 1.4; 95% CI 0.92–2.07) (Table 5.11). Symptom severity was positively associated with absence, and there was a dose-response relationship between symptom severity and the odds of absence (Figure 5.5). Those with severe compared to mild

symptoms had five times the odds of absence (OR 5.3; 95% CI 3.54–7.93). There was also an inverse relationship between age and absence; as participants aged their odds of absence decreased, however this association was weak in strength (OR 0.99; 95% CI 0.98–1.00).

Figure 5.5 Attenuated association between NS-SEC and sickness absence



Assumptions

The appropriateness of combining cases from the IID2 component studies was supported by analyses indicating the relationship between NS-SEC and the sickness absence outcome was not statistically significantly different between the Cohort and GP Presentation studies (Appendix 5).

Sensitivity analyses

Similar results to those reported were observed when analyses were conducted with recurrent episodes of IID included with clustering at the individual level accounted for using mixed-effects models (Appendix 5). Results from stratified analyses by age group also confirmed those from the main analysis (Appendix 5). Children (aged ≥5 to <16 years) and adults of

working age whose main household earner was in a routine/manual compared to managerial/professional occupation, had greater odds of experiencing sickness absence due to IID (OR 3.19; 95% CI 1.13–11.48 for children; and OR 1.48; 95%CI 1.04–2.12 for adults). This finding was also apparent when cases of all ages (0 to 90+ years) were analysed, however the association did not reach statistical significance (OR 1.20; 95% 0.95–1.52), which may have been related to measurement error when assessing sickness absence in young children aged <5 years. Increasing symptom severity was associated with greater odds of sickness absence for children, adults and those of all ages combined (Appendix 5).

Results from the multiply imputed datasets (detailed in Appendix 5) confirmed those from this analysis using listwise deletion, however the magnitude of the association between sickness absence and NS-SEC was weaker when the multiply imputed datasets were analysed. For cases in routine/manual compared to managerial/professional occupations the odds of sickness absence due to IID were 1.38 (95% CI 1.00–1.89) when using the multiply imputed dataset for cases of school or working age. Again, this association was attenuated and rendered non-significant when symptom severity was accounted for.

Additionally, comparable associations were found when investigating predictive factors for the duration of absence among absentees (Appendix 5).

5.8 Discussion

This analysis of data obtained from the largest population-based survey of IID conducted in the UK, found that individuals with IID of lower SES compared to high were more likely to experience severe symptoms, and were more likely to be absent from work, school or their normal daily activities. The association between SES and sickness absence was largely explained by greater symptom severity amongst the more disadvantaged groups.

These findings are comparable to those of previous studies that have analysed measures of IID severity and SES, however these other studies are sparse in number, and have focused on paediatric populations only. The findings presented here suggest that the association between SES and IID severity is true for the whole (all age) population, not just for children. One previous British study analysed data from the population-based Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC), and investigated predictive factors for the duration of

diarrhoeal episodes in children aged <6 months (Baker, Taylor and Henderson, 1998). The authors found an association between housing tenure and duration of diarrhoea, with infants living in rented versus mortgaged/owned accommodation (a suggested indicator of SES) having greater odds of experiencing diarrhoea for six or more days (OR 1.34; 95% CI 1.03–1.75). However, this association was attenuated and rendered non-significant after adjustment for duration of breast feeding, with longer spells of breast feeding providing protection against persistent diarrhoea.

Whilst very few cases were admitted to hospital in the present sample (<1%), the findings observed are somewhat similar to those of studies conducted in hospital settings. At this severe end of the disease spectrum, one UK-based study found low SES (measured by occupational social class) was associated with longer time to discharge for children aged <16 years hospitalised with gastroenteritis in univariate analysis (Conway, Phillips and Panday, 1990). Similarly, among American children aged <5 years hospitalised with gastroenteritis, those enrolled in Medicaid (a proxy measure for low SES) experienced longer average length of stay, compared to children not enrolled, when no other factors were taken into consideration (Ma, El Khoury and Itzler, 2009). In contrast, multivariate analysis revealed that education level and income were not related to length of stay for Canadian children aged <5 years, hospitalised with rotavirus gastroenteritis, whereas regularly seeing a physician for a medical condition was associated with longer hospital stays (Ford-Jones et al., 2000).

Together, these findings might suggest that the association between SES and IID severity could be mediated by socially patterned factors that impair immune response, such as lack of breast feeding in infancy and multimorbidity (Jackson and Nazar, 2006; Castle et al., 2005), both of which are more prevalent among lower socioeconomic groups (Oakley et al., 2013; Barnett et al., 2012). Additional biologically plausible mechanisms which might help to explain a greater burden of severe IID in lower socioeconomic groups, but are as yet to be substantiated in this context, include increased levels of chronic stress, smoking and nutritional deficiencies, all of which display social gradients and are associated with immune system compromise (Lund and O'Brien, 2011; Stämpfli and Anderson, 2009; Darmon and Drewnowski, 2008; Cohen, Doyle and Baum, 2006; Segerstrom and Miller, 2004). The potential mediating role of immune suppressing variables on the relationship between SES and symptom severity warrants further investigation.

I found that individuals with IID of lower SES compared to high had greater odds of sickness absence due to IID, and this was largely explained by greater symptom severity amongst IID cases of lower SES. In a large cohort of UK civil servants, age adjusted rates of

sickness absence due to gastroenteritis lasting seven days or less, were over six and four times higher for men and women respectively, in lower employment grades compared to high (Feeney et al., 1998). A social gradient was also observed for absences lasting longer than seven days. Conversely, self-reported sickness absence for gastroenteritis in a cohort of Dutch employees was unrelated to education level in univariate analysis (Mohren et al., 2005). These inconsistent findings may, in part, be due to different characteristics of the populations studied, since the age, sex and ethnicity adjusted results for absence presented in this thesis, were akin to those observed in the UK-based study of civil servants (Feeney et al., 1998). However, neither study investigated the role of symptom severity, which was identified as an important potential mediator of the relationship between SES and sickness absence due to IID in my analysis.

There are several limitations to this analysis. The validity of the results depended upon the unbiased and accurate self-reporting of symptoms and sickness absence among cases. If those of lower SES perceived their symptoms differently to those of higher SES, which has been observed in studies investigating perceptions of pain across socioeconomic groups (Miljković et al., 2014; Dorner et al., 2011), the results could be a mere artefact of the severity measurement. Nonetheless, the variables used to derive the symptom severity score in this study were related to the presence and duration of symptoms, which are rather more objective measures of severity compared to, for example, a subjective rating of symptom severity from mild to severe.

There was a large amount of missing data, particularly within the NS-SEC and symptom severity variables. Listwise deletion as a method of handling missing data can produce unbiased estimates when data are missing completely at random (Kang, 2013). However, the odds of whether data were missing or not within the NS-SEC and symptom severity variables were associated with other variables within the dataset (Appendix 5), supporting the idea that missing data were missing at random, rather than missing completely at random. Sensitivity analyses were therefore performed using multiple imputation by chained equations to impute missing data values. Results from multiply imputed datasets confirmed those from the main analyses (Appendix 5), suggesting that any bias resulting from the use of listwise deletion, was minimal. Ethnicity however was not associated with symptom severity when analyses were performed using the imputed datasets.

Cases identified in the IID2 Cohort and GP Presentation studies were combined for this analysis. Individuals in managerial/professional occupations, those aged 55+ years and those of White ethnicity were over-represented in the Cohort study compared to the UK

population, and individuals in intermediate and routine/manual occupations and those aged 15–24 years in particular were under-represented (Tam et al., 2012b). Under-representation of lower socioeconomic groups is commonplace in population-based surveys (Lorant et al., 2007), and could limit the external validity of the findings presented. Nevertheless, the internal validity of the findings should remain unaffected. It is possible that if non-participation or the design of the studies resulted in the under-representation of cases of lower SES who experienced milder symptoms, then the association between low SES and severe symptoms may have been overestimated. However, within the Cohort study this seems unlikely as cases were captured prospectively. The GP Presentation study may have been more prone to selection bias, since cases with more severe symptoms and those of lower SES may be more likely to present to their GP for an episode of IID (Tam, Rodrigues and O'Brien, 2003). However, as shown in Appendix 5, the relationships between NS-SEC and symptom severity, and NS-SEC and sickness absence, were not statistically significantly different between the Cohort and GP Presentation studies.

There is the potential for different pathogens to infect people of different SES, for example norovirus and *Listeria* have been associated with low SES in some studies (Phillips et al., 2011; Gillespie et al., 2010). Unfortunately, I was unable to explore the role of pathogen type on the association between SES and symptom severity because for around 58% of the sampled cases no pathogen was identified (Tam et al., 2012b). The impact of pathogen type on the association between SES and symptom severity is unknown, however the severity of illness likely depends not only on the infecting pathogen but also on host factors and the dose to which the host is exposed (O'Brien and Halder, 2007). The relationship between SES, pathogen type and IID symptom severity could be explored using a larger sample of cases, since for the majority a pathogen will not be identified.

Finally, the IID2 study also contained a retrospective telephone survey which gave higher IID incidence estimates compared to the IID2 Cohort study (Tam et al., 2012b), however I was unable to repeat the analyses with cases identified in the telephone survey because NS-SEC information was not collected. I was also unable to assess inequalities in sickness absence amongst those providing care for IID cases (caregiver informative was not collected) however this may be an interesting avenue for further research.

This study sheds new light into an under-researched area and indicates that the consequences of having an IID may be unequally shared across socioeconomic groups. These consequences are potentially serious. Loss of working days due to sickness can have important economic consequences and these are likely to be more severe for more

disadvantaged groups who might receive less adequate compensation from their employer. Loss of days from school can affect educational attainment (Department for Education, 2015), suggesting that the unequal effects of IID could exacerbate educational inequalities. Actions that reduce the risk of acquiring IID are unlikely to sufficiently address these inequalities; public health interventions also need to reduce their unequal consequences. Further research is required to understand the mechanisms explaining greater severity of illness in disadvantaged groups, and to identify ways to minimise the differential impact of IID on sickness absence.

Chapter 6

Results: Study 3

Impact of socioeconomic inequalities and neighbourhood characteristics on emergency hospitalisations for IID in England

6.1 ABSTRACT

Background

This study was designed to address gaps in the literature identified in Study 1, and to further investigate and build upon the findings of Study 2 of this thesis. In this final study, inequalities in IID-related emergency hospital admission rates for adults, as well as children were investigated. Furthermore, inequalities in the duration of emergency hospital admissions for IID were assessed, to gain further insight into the social patterning of the severity of IID.

Methods

A cross-sectional, ecological analysis was performed using HES data on emergency hospital admissions and admission durations for IID, collected over the period from 2009 to 2015 across England. Data analysis was conducted at the neighbourhood/LSOA level for three age groups. Data linkage techniques and regression modeling were used to assess the relationship between neighbourhood income deprivation and emergency admission rates for IID and admission durations, whilst controlling for the effects of several neighbourhood-level characteristics.

Results

Amongst over 30,000 neighbourhoods, increasing income deprivation was statistically significantly associated with increasing emergency hospital admission rates for IID and increasing duration of admissions, for all ages. The associations between deprivation and admission rates for IID were attenuated but remained statistically significant after controlling for factors such as the higher prevalence of long-term health problems in the more deprived neighbourhoods, and the closer proximity of the more deprived neighbourhoods to hospitals with A&E departments. The prevalence of long-term health problems appeared to be an important mediator of the association between deprivation and IID admission duration for adults aged 15–64 years, but not for children or older adults aged 65+ years.

Conclusions

This study found socio-spatial gradients in IID-related emergency hospital admission rates and admission durations, which appear to exist not only for children but for adults as well. Reducing IID hospital admission rates and admission durations experienced by the most disadvantaged to levels experienced by the least, should be an important goal for any levelling-up policy or intervention designed to improve equity in health.

6.2 INTRODUCTION

The final study of this thesis has been designed to address gaps in the literature identified in Study 1, and to further investigate and build upon the findings of Study 2. To introduce this last study, it is therefore necessary to reflect back on the results presented in previous chapters. The first study in this thesis (Study 1), a systematic literature review, aimed to investigate the relationship between SES and GI infection risk in high income countries. Age was identified as a statistically significant potential effect modifier of the relationship between SES and GI infection risk, however the majority of studies that assessed inequalities in hospital admissions for IID also analysed children, and very few investigated inequalities in IID-related hospital admissions amongst adults specifically. Additionally, Study 2 of this thesis (an analysis of the population-based IID2 study) found that individuals with IID of lower SES compared to high were more likely to experience severe symptoms, a finding that was observed for adults as well as children. A hypothesis was generated that the association between SES and IID severity might be mediated by socially patterned factors that impair immune response, such as comorbidity.

This study aims to investigate socioeconomic inequalities in emergency hospital admission rates for IID in England, and to assess inequalities in the duration of these admissions. I investigate inequalities in admission rates for IID for adults aged 15–64 years and older adults aged 65+ years, as well as children. In doing this, I hope to address a gap in the literature identified in Study 1. Furthermore, by investigating inequalities in the duration of admissions for IID, I hope to gain further insight into the social patterning of the severity of IID, and thus build upon the findings of Study 2. Additionally, I investigate a hypothesis generated from Study 2, by examining the potential mediating role of long-term health problems on inequalities in admission durations for IID.

In the UK, three previous studies have investigated inequalities in IID-related hospital admission rates, and one study has examined inequalities in admission duration for acute gastroenteritis. Firstly, Kyle et al. (2011) and Pockett et al. (2011) conducted ecological analyses to assess the relationship between neighbourhood deprivation and hospital admission rates due to IID in children (aged 0–14 and <5 years, respectively), aggregating data over large areas (PCT-level). Kyle et al. (2011) analysed admissions in the Greater London region of England and found no statistically significant correlations between admission rates and deprivation, whilst Pockett et al. (2011) analysed data across the whole of England and found admission rates statistically significantly increased with increasing deprivation. Additionally, Olowokure et al. (1999) examined inequalities in admissions for

IID across all ages, and found admission rates increased with increasing deprivation across five age groups (ranging from 0 to >75 years). However, the data used in the analysis were collected over twenty years ago (years 1990–5) and were limited to the West Midlands region in England. In relation to admission duration (length of stay), Conway, Phillips and Panday (1990) studied children aged <16 years admitted to a UK-based hospital infectious disease unit for acute gastroenteritis. They found low SES was associated with longer time to discharge for children hospitalised with gastroenteritis, however data were collected in the 1980s. None of the studies controlled for potential confounding variables, although Olowokure et al. (1999) stratified results by age.

An additional study by Busby, Purdy and Hollingworth (2017b) found that admission rates for dehydration and gastroenteritis were higher among the most deprived GPs in England compared to the least, adjusting for practice characteristics, such as access to care, continuity of care and distance to the nearest hospital. However, individuals with specified non-infectious gastroenteritis (e.g. allergic and dietetic gastroenteritis) were included in the case definition. Admission data were collected over one year (2011–12), and the analysis was performed for all ages combined.

The study presented in this chapter is unique and adds to the knowledge base by providing up-to-date measures of the associations between deprivation and IID-related emergency hospital admission rates, and admission durations, for all English neighbourhoods, over the period from 2009 to 2015. Moreover, small area measures are used to aggregate data, and results are stratified by age which permits the comparison of inequalities in admission rates and the duration of admissions between adults and children. Using data linkage techniques, I also investigate the potential effects of several neighbourhood-level characteristics on the relationship between deprivation and admission rates and admission duration for IID.

6.3 AIM AND OBJECTIVES

Aim

- To assess the impact of neighbourhood income deprivation on emergency hospitalisations for IID in England using HES data

Objectives

- To investigate the relationships between neighbourhood income deprivation and:
  - Emergency hospital admission rates for IID
  - Admission days per emergency admission for IID
- To explore the potential effects of several neighbourhood-level characteristics on the relationship between SES and the outcomes above

6.4 METHODS

The methods are described in detail in Chapter 3.

Ethical considerations

For this analysis, a Data Access Request to the ILRR at the University of Liverpool was made to access anonymised and aggregated HES datasets. Cells with small counts (n<5) that also had small underlying pooled population sizes (n<1000) were suppressed as required by the Anonymisation Standard for Publishing Health and Social Care Data Specification (Information Standards Board for Health and Social Care, 2013) before being released to me, thus minimising any risk of re-identification.

6.5 RESULTS

Descriptive statistics

Characteristics of the LSOAs are displayed in Table 6.1, stratified by income deprivation and three age groups. For all three age groups, the LSOAs that were most deprived compared to the least, had higher crude emergency hospital admission rates for IID, and higher numbers of admission days per admission. For adults aged 65+ years, emergency hospital admission rates in the most deprived LSOAs were more than twice the rates in the least deprived. More

deprived LSOAs had greater proportions of residents of Non-White ethnicity, and residents with long-term health problems or disabilities. A greater proportion of the most deprived LSOAs compared to the least, were classified as urban, and more deprived LSOAs tended to be situated closer to GP surgeries and hospitals with A&E departments.

Overall, children and adults aged 65+ years had higher emergency hospital admission rates for IID (4.76 and 4.92 admissions per 1000 population per year, respectively) compared to adults aged 15–65 years (1.1 admissions per 1000 population per year). However, on average children spent less time in hospital per admission compared to adults aged 15–65 years, and adults aged 65+ years (1, 2.89 and 8.67 days per admission, respectively). The proportion of those with a long-term health problem or disability, and the proportion of those of White ethnicity, increased as the age groups increased from children to older adults (Table 6.1).

Correlation matrices are displayed in Figures 6.1 to 6.3, which graphically demonstrate correlations between the independent variables. For adults aged 15–64 years and adults aged 65+ years, income deprivation and the proportion of the population with a long-term health problem or disability, were highly positively correlated (Pearson *r* correlation coefficient: 0.81 and 0.72, respectively).

Missing data

For adults aged 15–65 years, all of the 32,844 LSOAs in England were available to analyse. Due to the suppression of small numbers, only 32,212 (98%) and 30,502 (93%) LSOAs in England were available to analyse for children and adults aged 65+ years, respectively. It was considered that this small amount of missing data was unlikely to substantively influence the results.

A very small number of LSOAs had no admissions for IID (children [n=399, 1.2%]; adults aged 15–64 years [n=232, 0.7%]; adults aged 65+ years [n=216, 0.7%]). These LSOAs were excluded when analysing the admission days per admission outcome, since they were deemed not applicable to the analysis.

Table 6.1 Characteristics of English LSOAs stratified by income deprivation quintiles, for three age groups

	Children aged 0–14 years							Adults aged 15–64 years							Adults aged 65+ years						
	Income deprivation							Income deprivation							Income deprivation						
	Q1	Q2	Q3	Q4	Q5	p-value[a]	All	Q1	Q2	Q3	Q4	Q5	p-value[a]	All	Q1	Q2	Q3	Q4	Q5	p-value[a]	All
Number LSOAs	6614	6373	6363	6450	6412		32212	6606	6613	6560	6531	6534		32844	6296	5914	6122	6101	6069		30502
IID admission rate per 1000 pop per year mean (SD)	3.39 (2.59)	4.05 (2.57)	4.68 (2.78)	5.42 (3.18)	6.29 (3.40)	<0.001	4.76 (3.09)	0.78 (0.42)	0.90 (0.45)	1.03 (0.48)	1.23 (0.54)	1.55 (0.66)	<0.001	1.10 (0.58)	3.34 (1.84)	3.97 (2.09)	4.63 (2.33)	5.64 (2.81)	7.06 (3.42)	<0.001	4.92 (2.88)
IID admission days per admission [b] mean (SD)	0.98 (0.90)	0.98 (0.94)	0.99 (0.94)	1.03 (0.92)	1.03 (0.76)	<0.001	1.00 (0.89)	2.75 (2.58)	2.84 (2.50)	2.90 (2.46)	2.96 (2.14)	2.99 (2.01)	<0.001	2.89 (2.35)	8.44 (6.85)	8.69 (6.07)	8.71 (6.16)	8.82 (6.43)	8.72 (5.93)	0.018	8.67 (6.30)
Percentage White ethnicity mean (SD)	90.10 (10.44)	88.79 (13.66)	82.16 (20.60)	74.68 (26.41)	67.69 (30.25)	<0.001	80.72 (23.18)	92.86 (8.16)	92.45 (10.13)	87.80 (16.11)	82.18 (21.05)	76.19 (25.01)	<0.001	86.33 (18.41)	97.67 (3.84)	97.18 (5.85)	94.61 (11.28)	91.43 (15.36)	88.11 (18.88)	<0.001	93.81 (12.89)
Percentage long-term health problem mean (SD)	2.59 (1.22)	3.07 (1.32)	3.50 (1.43)	4.04 (1.52)	4.68 (1.58)	<0.001	3.57 (1.60)	8.12 (2.27)	10.07 (2.34)	11.67 (2.63)	14.09 (3.11)	18.70 (4.12)	<0.001	12.51 (4.70)	40.88 (6.61)	45.42 (6.61)	49.87 (7.15)	55.25 (7.88)	62.64 (8.36)	<0.001	50.77 (10.59)
Distance to GP in km mean (SD)	2.00 (1.60)	1.96 (1.85)	1.46 (1.51)	1.06 (1.27)	0.88 (0.64)	<0.001	1.47 (1.50)	1.97 (1.59)	1.95 (1.85)	1.47 (1.53)	1.06 (1.27)	0.88 (0.64)	<0.001	1.47 (1.50)	2.03 (1.63)	1.99 (1.88)	1.52 (1.58)	1.09 (1.31)	0.89 (0.65)	<0.001	1.51 (1.54)
Distance to hospital in km mean (SD)	11.83 (8.48)	12.43 (9.55)	11.34 (9.92)	8.98 (8.30)	6.73 (6.20)	<0.001	10.26 (8.84)	11.65 (8.47)	12.40 (9.58)	11.35 (9.95)	9.04 (8.39)	6.75 (6.23)	<0.001	10.25 (8.87)	11.92 (8.51)	12.62 (9.62)	11.66 (10.03)	9.40 (8.62)	6.89 (6.32)	<0.001	10.50 (8.96)
Rural %	27	28.9	18.5	8.6	2.4	<0.001	17.1	26.1	29	18.7	8.7	2.4	<0.001	17	28	30.1	20.2	9.6	2.7	<0.001	18.1

GP = general practice; IID = infectious intestinal disease; km = kilometres; LSOA = Lower Super Output Area; pop = population; SD = standard deviation

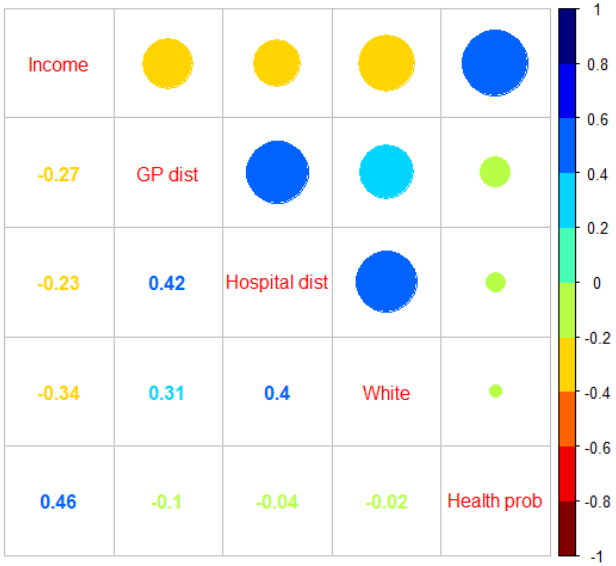
Q1 = least income deprived quintile; Q5 = most income deprived quintile

[a] Statistical significance of relationship between income deprivation and each variable, tested using one-way analysis of variance (ANOVA) and χ^2 test for rural/urban classification

[b] Not applicable LSOAs for admission days per admission (LSOAs with 0 admissions): children aged 0–14 years (number=399, 1.2%); adults aged 15–64 years (number=232, 0.7%); adults aged 65+ (number=216, 0.7%)

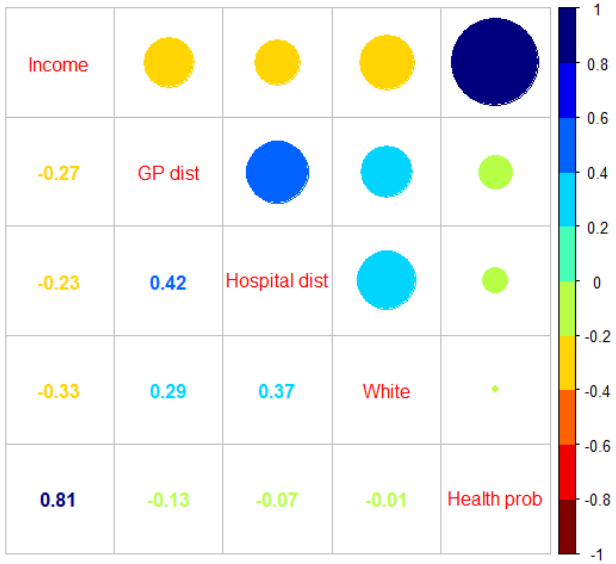
Figures 6.1 to 6.3 Correlation matrices showing relationship between independent variables for children, adults and older adults

Figure 6.1 Children aged 0–14 years



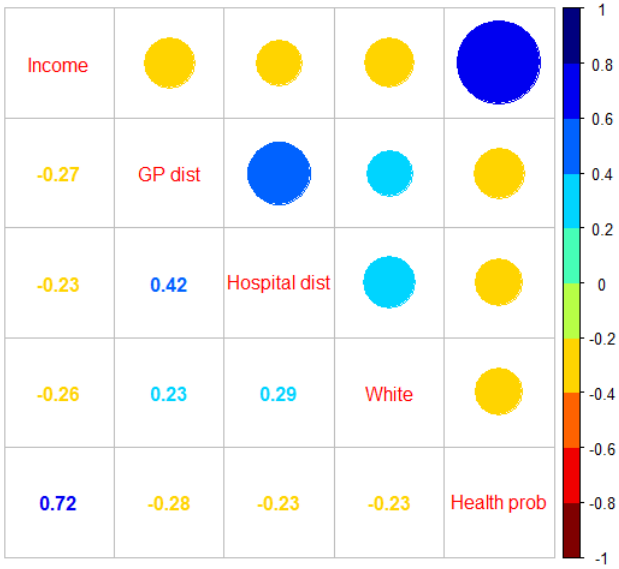
Number LSOAs = 32212

Figure 6.2 Adults aged 15–64 years



Number LSOAs = 32844

Figure 6.3 Adults aged 65+ years



Number LSOAs = 30502

Legend:
Income = income deprivation; GP dist = average distance to nearest GP in kilometres; Hospital dist = average distance to nearest hospital with A&E department in kilometres; White = proportion of the population of White ethnicity; Health prob = proportion of the population with a long-term health problem or disability.

Statistics displayed within the plots are Pearson *r* correlation coefficients. Larger circles and darker colours represent higher correlations between the variables (colour key displayed on right-hand side of plots).

Univariate analysis

Results of the univariate analyses for the two hospitalisation outcomes are displayed in Tables 6.2 and 6.3. Increasing LSOA/neighbourhood income deprivation was statistically significantly associated with increasing IID-related emergency hospital admission rates and log admission days per admission, for all age groups. The most deprived neighbourhoods compared to the least had approximately two times the rate of emergency hospital admissions for IID, and there was a trend across the five deprivation quintiles of increasing admission rates with increasing deprivation in each age group. The magnitude of this association was strongest for older adults aged 65+ years, followed by adults aged 15–64 years and children (IRR: 2.13 [95% CI 2.09–2.16]; 1.97 [95% CI 1.94–2] and 1.88 [95% CI 1.84–1.91], respectively). In comparison, the magnitude of the association between deprivation and log admission days per admission for IID, was strongest for adults aged 15–64 years, followed by children and older adults aged 65+ years. In terms of percent change, for every additional 10% of the population experiencing deprivation relating to low income, a 6.09%, 4.16% and 2.6% increase in admission days per admission for IID would be expected for adults, children and older adults, respectively.

The proportion of the population with a long-term health problem or disability was also statistically significantly positively associated with IID-related emergency hospital admission rates and log admission days per admission, for all age groups. For both outcomes, the magnitude of this association was strongest for adults aged 15–64 years. Geographical variables such as rural compared to urban neighbourhoods, and neighbourhoods situated further away from GPs and hospitals with A&E departments, were statistically significantly associated with lower emergency hospital admission rates for IID (Table 6.2). Rural neighbourhoods and those situated further away from health services were also associated with shorter admission durations for children. However, associations between the geographical variables and the log admission days per admission outcome were less apparent for adults (Table 6.3).

Table 6.2 Univariate negative binomial regression models for IID emergency hospital admission rates for English LSOAs, stratified by age

	IID emergency hospital admission rates IRR (95% CI)					
	Children aged 0–14 years		**Adults aged 15–64 years**		**Adults aged 65+ years**	
Proportion White (≤70%)	ref		ref		ref	
>70% to ≤90%	0.97	(0.95 -0.99)*	0.87	(0.86 -0.89)***	0.89	(0.86 -0.92)***
>90%	1.04	(1.02 -1.06)*	0.87	(0.86 -0.89)***	0.68	(0.66 -0.70)***
Income deprivation (Q1)	ref		ref		ref	
Q2	1.20	(1.17 -1.23)***	1.15	(1.13 -1.17)***	1.19	(1.16 -1.21)***
Q3	1.39	(1.36 -1.42)***	1.31	(1.29 -1.33)***	1.38	(1.36 -1.41)***
Q4	1.61	(1.58 -1.65)***	1.56	(1.54 -1.59)***	1.69	(1.66 -1.72)***
Q5 (most deprived)	1.88	(1.84 -1.91)***	1.97	(1.94 -2.00)***	2.13	(2.09 -2.16)***
Proportion long-term health problem (Q1)	ref		ref		ref	
Q2	1.18	(1.16 -1.20)***	1.23	(1.21 -1.25)***	1.17	(1.15 -1.19)***
Q3	1.30	(1.27 -1.32)***	1.46	(1.44 -1.48)***	1.40	(1.38 -1.43)***
Q4 (greatest proportion)	1.44	(1.41 -1.47)***	1.91	(1.89 -1.94)***	1.86	(1.83 -1.89)***
Distance GP (km)	0.96	(0.95 -0.96)***	0.94	(0.94 -0.95)***	0.90	(0.89 -0.90)***
Distance hospital (km)	0.99	(0.99 -0.99)***	0.99	(0.99 -0.99)***	0.98	(0.98 -0.98)***
Classification (Urban)	ref		ref		ref	
Rural	0.81	(0.79 -0.83)***	0.79	(0.78 -0.80)***	0.66	(0.65 -0.67)***
Number of LSOAs	**32212**		**32844**		**30502**	

CI = confidence interval; GP = general practice; IID = infectious intestinal disease; IRR = incident rate ratio; km = kilometre; LSOA = Lower Super Output Area; ref = reference category
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

Table 6.3 Univariate linear regression models for log IID admission days per emergency admission for English LSOAs, stratified by age

	Log IID admission days per admission‡		
	Children aged 0–14 years	**Adults aged 15–64 years**	**Adults aged 65+ years**
Proportion White (%)†	-0.0399***	0.0007	0.0367***
std. error	0.0013	0.0019	0.0030
Income deprivation (%)†	0.0408***	0.0591***	0.0257**
std. error	0.003	0.0034	0.0039
Proportion long-term health problem (%)†	0.0386*	0.1608***	0.0174*
std. error	0.0194	0.0075	0.0037
Distance GP (km)	-0.0289***	-0.0035	-0.0084*
std. error	0.0021	0.0024	0.0026
Distance hospital (km)	-0.0058***	-0.0005	0.0006
std. error	0.0003	0.0004	0.0004
Rural [a]	-0.1091***	-0.0201*	-0.0026
std. error	0.0082	0.0095	0.0102
Number of LSOAs	**31813**	**32612**	**30286**

GP = general practice; IID = infectious intestinal disease; km = kilometre; LSOA = Lower Super Output Area; std. error = standard error
[a] Reference category = Urban
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}
‡ Regression coefficients and standard errors displayed in table
† Variables entered into model in units of 10% points

Multivariate analysis

Emergency hospital admission rates for IID

In multivariate analysis, increasing neighbourhood income deprivation was statistically significantly associated with increasing emergency hospital admission rates for IID, across all of the models, for all age groups (Tables 6.4, 6.5 and 6.6). In general, the magnitude of this association was attenuated following adjustment for the proportion of the population with a long-term health problem or disability, and the geographical variables, but remained statistically significant. Across the deprivation quintiles admission rates increased with increasing deprivation in each age group. In the fully adjusted model (Model 4), children

living in the most deprived neighbourhoods compared to the least had nearly two times the rate of emergency hospital admissions for IID (IRR 1.89; 95% CI 1.84–1.93). Older adults aged 65+ years and adults aged 15–64 years living in the most deprived neighbourhoods compared to the least had 1.7 and 1.4 times the rate of admissions for IID respectively, in the fully adjusted models (IRR: 1.67 [95% CI 1.64–1.71] and 1.43 [95% CI 1.40–1.47], respectively).

The proportion of the population with a long-term health problem or disability was also statistically significantly positively associated with emergency hospital admission rates for IID, for all age groups. The magnitude of this association was strongest for adults aged 15–64 years (Model 4: largest compared to smallest proportion of the population with a long-term health problem IRR 1.46; 95% CI 1.42–1.49). Neighbourhoods situated further away from A&E hospitals and rural compared to urban neighbourhoods had statistically significantly lower emergency hospital admission rates for IID, and the magnitudes of these associations were similar across the age groups. Neighbourhoods where more than 90% of the child residents were of White ethnicity compared to ≤70%, had statistically significantly higher IID emergency hospital admission rates for children (Model 4 IRR: 1.38; 95% CI 1.35–1.40), whereas the direction of this association was reversed for adults aged 65+ years (Model 4 IRR: 0.95; 95% CI 0.93–0.97).

Admission days per admission for IID

Multivariate analysis for the IID admission duration outcome (Table 6.7) showed that increasing neighbourhood income deprivation was statistically significantly associated with increasing log admission days per admission for IID, in models controlling for ethnicity, for all age groups. The magnitude of this association appeared to be strongest for adults aged 15–64 years, followed by older adults aged 65+ years and children. In terms of percent change, the results from the models controlling for ethnicity (Model 2) indicate that for every additional 10% of the population experiencing deprivation relating to low income, a 1.17%, 6.88% and 4.17% increase in admission days per admission for IID would be expected for children, adults and older adults, respectively.

For adults aged 65+ years, the association between deprivation and log admission days per admission was attenuated following adjustment for the proportion of the population with long-term health problems, and the geographical variables, but remained statistically significant. For children, the association between deprivation and log admission days per

admission was slightly stronger following adjustment for the proportion of the population with long-term health problems, but was attenuated after controlling for the geographical variables (but remained statistically significant). However, for adults aged 15–64 years, the statistically significant association between deprivation and log admission days per admission was attenuated and rendered non-significant following adjustment for long-term health problems.

The proportion of adults aged 15–64 years with a long-term health problem or disability was statistically significantly positively associated with log admission days per admission for adults aged 15–64 years, and this was the only variable that remained statistically significant in the fully adjusted model (Model 4). On the other hand, long-term health problems were not statistically significantly associated with admission duration for children or adults aged 65+ years. Additionally, rurality was not significantly associated with admission duration for IID, however neighbourhoods situated further away from GPs were associated with statistically significantly shorter admission durations for children and adults aged 65+ years. Neighbourhoods that had a greater proportion of children of White ethnicity were associated with statistically significantly shorter admission durations for children. The direction of this association was reversed for older adults, whereby neighbourhoods that had a greater proportion of adults aged 65+ years of White ethnicity were associated with statistically significantly longer admission durations for adults aged 65+ years.

The results from the fully adjusted model (Model 4) indicate that for every additional 10% of the population experiencing deprivation relating to low income, a 0.86% increase in admission days per admission would be expected for children, and a 3.69% increase in admission days per admission for IID would be expected for adults aged 65+ years, holding all other variables in the model constant.

Table 6.4 Multivariate negative binomial regression models for IID emergency hospital admission rates for English LSOAs, for children aged 0–14 years

	IID emergency hospital admission rates for children aged 0–14 years IRR (95% CI)			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Proportion White (≤70%)	ref	ref	ref	ref
>70% to ≤90%	0.97 (0.95 -0.99)*	1.13 (1.11 -1.15)***	1.12 (1.10 -1.15)***	1.14 (1.12 -1.16)***
>90%	1.04 (1.02 -1.06)*	1.27 (1.25 -1.29)***	1.26 (1.24 -1.28)***	1.38 (1.35 -1.40)***
Income deprivation (Q1)		ref	ref	ref
Q2		1.21 (1.18 -1.23)***	1.19 (1.17 -1.22)***	1.20 (1.17 -1.23)***
Q3		1.44 (1.41 -1.48)***	1.41 (1.38 -1.44)***	1.42 (1.39 -1.45)***
Q4		1.70 (1.67 -1.74)***	1.65 (1.62 -1.69)***	1.63 (1.59 -1.66)***
Q5 (most deprived)		2.03 (1.99 -2.07)***	1.95 (1.90 -2.00)***	1.89 (1.84 -1.93)***
Proportion long-term health problem (Q1)			ref	ref
Q2			1.07 (1.05 -1.09)**	1.07 (1.05 -1.09)***
Q3			1.09 (1.07 -1.11)***	1.09 (1.07 -1.11)***
Q4 (greatest proportion)			1.09 (1.06 -1.11)***	1.09 (1.06 -1.11)***
Distance GP (km)				1.01 (1.00 -1.01)
Distance hospital (km)				0.99 (0.99 -0.99)***
Classification (Urban)				ref
Rural				0.94 (0.92 -0.96)**
Log-likelihood	-99257.3	-96889.7	-96847.1	-96490.4
AIC	198522.6	193795.4	193716.2	193008.8
BIC	198556.1	193862.4	193808.3	193126.1
Number LSOAs	32212	32212	32212	32212

CI = confidence interval; GP = general practice; IID = infectious intestinal disease; IRR = incident rate ratio; km = kilometre; LSOA = Lower Super Output Area; ref = reference category
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

Table 6.5 Multivariate negative binomial regression models for IID emergency hospital admission rates for English LSOAs, for adults aged 15–64 years

	IID emergency hospital admission rates for adults aged 15–64 years IRR (95% CI)			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Proportion White (≤70%)	ref	ref	ref	ref
>70% to ≤90%	0.87 (0.86 -0.89)***	1.03 (1.01 -1.05)*	1.00 (0.98 -1.02)	1.01 (0.99 -1.03)
>90%	0.87 (0.86 -0.89)***	1.09 (1.08 -1.11)***	1.00 (0.98 -1.01)	1.06 (1.05 -1.08)***
Income deprivation (Q1)		ref	ref	ref
Q2		1.15 (1.13 -1.17)***	1.07 (1.05 -1.09)***	1.08 (1.06 -1.10)***
Q3		1.33 (1.30 -1.35)***	1.16 (1.14 -1.18)***	1.16 (1.14 -1.18)***
Q4		1.60 (1.57 -1.62)***	1.28 (1.25 -1.31)***	1.26 (1.23 -1.29)***
Q5 (most deprived)		2.03 (1.99 -2.06)***	1.47 (1.43 -1.51)***	1.43 (1.40 -1.47)***
Proportion long-term health problem (Q1)			ref	ref
Q2			1.16 (1.14 -1.17)***	1.16 (1.14 -1.18)***
Q3			1.26 (1.24 -1.28)***	1.27 (1.25 -1.29)***
Q4 (greatest proportion)			1.45 (1.42 -1.49)***	1.46 (1.42 -1.49)***
Distance GP (km)				1.00 (1.00 -1.01)
Distance hospital (km)				0.99 (0.99 -0.99)***
Classification (Urban)				ref
Rural				0.98 (0.96 -0.99)*
Log-likelihood	-91247.2	-87391.7	-86876.7	-86488
AIC	182502.4	174799.5	173775.4	173003.9
BIC	182536	174866.7	173867.8	173121.5
Number LSOAs	32844	32844	32844	32844

CI = confidence interval; GP = general practice; IID = infectious intestinal disease; IRR = incident rate ratio; km = kilometre; LSOA = Lower Super Output Area; ref = reference category
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

Table 6.6 Multivariate negative binomial regression models for IID emergency hospital admission rates for English LSOAs, for adults aged 65+ years

	IID emergency hospital admission rates for adults aged 65+ years IRR (95% CI)			
	Model 1	**Model 2**	**Model 3**	**Model 4**
Proportion White (≤70%)	ref	ref	ref	ref
>70% to ≤90%	0.89 (0.86 -0.92)***	1.01 (0.98 -1.04)	1.03 (1.00 -1.06)*	1.02 (0.99 -1.05)
>90%	0.68 (0.66 -0.70)***	0.87 (0.85 -0.89)***	0.88 (0.86 -0.90)***	0.95 (0.93 -0.97)*
Income deprivation (Q1)		ref	ref	ref
Q2		1.18 (1.16 -1.20)***	1.14 (1.12 -1.16)***	1.17 (1.14 -1.19)***
Q3		1.37 (1.34 -1.39)***	1.27 (1.24 -1.29)***	1.28 (1.26 -1.31)***
Q4		1.65 (1.62 -1.68)***	1.45 (1.42 -1.48)***	1.44 (1.41 -1.47)***
Q5 (most deprived)		2.06 (2.02 -2.10)***	1.72 (1.68 -1.76)***	1.67 (1.64 -1.71)***
Proportion long-term health problem (Q1)			ref	ref
Q2			1.08 (1.06 -1.09)***	1.05 (1.04 -1.07)**
Q3			1.15 (1.13 -1.17)***	1.09 (1.07 -1.11)***
Q4 (greatest proportion)			1.29 (1.26 -1.32)***	1.21 (1.19 -1.24)***
Distance GP (km)				0.98 (0.97 -0.98)***
Distance hospital (km)				0.99 (0.99 -0.99)***
Classification (Urban)				ref
Rural				0.92 (0.90 -0.94)***
Log-likelihood	-89793.1	-86333.3	-86036.2	-85285.3
AIC	179594.3	172682.5	172094.3	170598.6
BIC	179627.6	172749.1	172185.9	170715.2
Number LSOAs	30502	30502	30502	30502

CI = confidence interval; GP = general practice; IID = infectious intestinal disease; IRR = incident rate ratio; km = kilometre; LSOA = Lower Super Output Area; ref = reference category
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

Table 6.7 Multivariate linear regression models for log IID admission days per admission for English LSOAs, stratified by age

	Log IID admission days per admission‡											
	Children aged 0–14 years				Adults aged 15–64 years				Adults aged 65+ years			
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Proportion White (%)†	-0.0399***	-0.0382***	-0.0381***	-0.0350***	0.0007	0.0128**	0.0010	0.0007	0.0367***	0.0453***	0.0454***	0.0467***
std. error	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.002	0.003	0.003	0.003	0.003
Income deprivation (%)†		0.0116*	0.0120*	0.0086*		0.0665***	0.0008	0.0019		0.0409***	0.0377**	0.0362**
std. error		0.003	0.004	0.004		0.004	0.007	0.007		0.004	0.006	0.006
Proportion long term health problem (%)†			-0.0057	-0.0063			0.1594***	0.1584***			0.0044	0.0026
std. error			0.022	0.022			0.014	0.014			0.005	0.005
Distance GP (km)				-0.0056*				0.0035				-0.0141*
std. error				0.003				0.003				0.003
Distance hospital (km)				-0.0013*				-0.0001				0.0003
std. error				0.000				0.000				0.001
Rural [a]				-0.0114				-0.0023				0.0210
std. error				0.011				0.012				0.013
Log-likelihood	-25811.3	-25804.5	-25804.4	-25788.3	-31711.4	-31542.6	-31481.3	-31480.6	-31419.8	-31367.1	-31366.8	-31357.2
AIC	51628.6	51617	51618.9	51592.7	63428.7	63093.1	62972.7	62977.1	62845.5	62742.3	62743.6	62730.3
BIC	51653.7	51650.4	51660.7	51659.6	63453.9	63126.7	63014.6	63044.3	62870.5	62775.6	62785.2	62796.9
Number LSOAs	31813	31813	31813	31813	32612	32612	32612	32612	30286	30286	30286	30286

GP = general practice; IID = infectious intestinal disease; km = kilometre; LSOA = Lower Super Output Area; std. error = standard error

[a] Reference category = Urban

*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

‡ Regression coefficients and standard errors displayed in table

† Variables entered into model in units of 10% points

Assumptions

Negative binomial regression is suitable for modeling over-dispersed count data. To test whether negative binomial regression was appropriate to model the IID emergency hospital admission outcome, multivariate models were created using Poisson regression and the fit of these models were compared to the negative binomial models using the likelihood ratio test (data not shown). The results were strongly suggestive that negative binomial regression was more appropriate to model the data compared to Poisson regression.

The assumptions of the linear regression models for the admission duration outcome were assessed graphically. The plots (displayed in Appendix 6) suggested that the residuals of the fully adjusted models (Model 4) were approximately normally distributed, and the variance of the residuals were approximately equal. This suggested that linear regression was appropriate for modeling the log admission days per admission outcome.

Finally, whilst each LSOA was entered into the models only once (i.e. there were no duplicates at the LSOA level), it was not possible to account for repeat admissions by the same person or control for any clustering at the hospital level since the data required to perform these checks were not available. Therefore, the assumption of independent observations may not have been met, which should be considered when evaluating the study findings.

Sensitivity analyses

To investigate the impact of including cases of unspecified non-infectious gastroenteritis (ICD-10 code K52.9) within the definition of IID, the analysis was repeated using a more specific definition of IID that excluded all cases of potentially non-infectious gastroenteritis of unspecified origin (ICD-10 codes K52.9 and A09.9).

Using this more specific definition of IID, emergency hospital admission rates were lower for all age groups, however similar to the main analysis, children and adults aged 65+ years had higher emergency hospital admission rates for IID (2.89 and 1.44 admissions per 1000 population per year, respectively) compared to adults aged 15–65 years (0.32 admissions per 1000 population per year). On average, children were admitted for 1.01 days per admission, which was similar to that observed in the main analysis, however admission durations for adults aged 15–64 and 65+ years were longer when calculated using the more specific

definition of IID compared to that used in the main analysis. Using the more specific definition, adults aged 15–64 and 65+ years were admitted for an average of 3.58 and 12.73 days per admission, respectively.

The results of the regression analysis are displayed in Appendix 6, and confirmed those of the main analysis. Inequalities in emergency hospital admission rates and admission duration for IID were observed for all age groups, indicating that the inclusion of cases of unspecified non-infectious gastroenteritis had a minimal impact on the main findings of the analysis.

6.6 Discussion

This cross-sectional ecological analysis of English HES data from 2009–15, found that increasing neighbourhood income deprivation was statistically significantly associated with increasing emergency hospital admission rates for IID and increasing duration of admissions for IID, for all ages. The associations between deprivation and emergency hospital admission rates for IID were partly explained by factors such as the higher prevalence of long-term health problems in the more deprived neighbourhoods, and the closer proximity of the more deprived neighbourhoods to hospitals with A&E departments. The prevalence of long-term health problems appeared to be an important mediator of the association between deprivation and IID admission duration for adults aged 15–64 years, but not for children or older adults aged 65+ years.

As highlighted in the introduction section above, three previous studies have investigated inequalities in IID-related hospital admission rates in the UK. Two of these studies observed socio-spatial gradients in IID-related hospital admissions, one of which analysed children only (Pockett et al., 2011), and the other analysed individuals of all ages in the West Midlands (Olowokure et al., 1999). One other study, found no statistically significant relationship between deprivation and hospital admission rates for diarrhoea amongst children living in London (Kyle et al., 2011). The study presented in this chapter, provides an up-to-date assessment of the extent of inequalities in IID emergency hospital admission rates for all ages, and confirms the results of Pockett et al. (2011) and Olowokure et al. (1999). In England, a socio-spatial gradient in emergency hospital admission rates for IID was observed for children aged 0–14 years, adults aged 15–64 years and adults aged 65+ years.

Similar associations have been observed by studies conducted in other high income countries, although as highlighted in Study 1 of this thesis, these studies have tended to focus on paediatric populations. Studies conducted in countries such as Denmark, Germany, the USA, Turkey, Italy, New Zealand and Australia have observed inequalities in IID-related hospital admissions for children (Biering-Sørensen et al., 2012; Wilking et al., 2012; Ma, El Khoury and Itzler, 2009; Özmert, Kilic and Yurdakök, 2008; Dennehy et al., 2006; Borgnolo et al., 1996) and for all ages combined (Lal et al., 2012; Moorin et al., 2010). Whilst not all studies conducted in high income countries have observed statistically significant associations between SES and IID-related hospital admission rates (Xu, Hu and Tong, 2015; Seo et al., 2013; Teschke et al., 2010; Kum-Nji et al, 2009), the majority of evidence (including this present study) suggests a social gradient in hospital admissions for IID does exist.

Acute gastroenteritis is classified as an ACSC, meaning that hospital admission for this condition could be avoided through early intervention and effective management (Ham, Imison and Jennings, 2010). The findings from this study might therefore reflect inadequacies in the provision or quality of primary care in deprived neighbourhoods. In an analysis of 28 different ACSCs, Busby, Purdy and Hollingworth (2017b) found that factors indicative of primary care quality such as higher continuity of care were associated with lower rates of unplanned hospital admissions for dehydration and gastroenteritis (a definition that included specified non-infective gastroenteritis). However, questions remain as to whether improving continuity of care would lead to reductions in inequalities in emergency hospital admission rates for IID. This research area warrants further investigation.

The study presented in this chapter also found that increasing neighbourhood income deprivation was statistically significantly associated with increasing duration of hospital admission for IID, for all ages. Hospitalisation is in itself a severe consequence of IID, and could be viewed as a proxy measure of IID severity, however inequalities in hospital admission rates might also reflect differential risk of infection by SES. Therefore it was considered that admission duration might be a more accurate measure of disease severity compared to hospital admission rates, with the caveat that admission duration may also be influenced by non-medical factors such as issues with social care. The findings of this study confirm those of Study 2 of this thesis, which found inequalities in symptom severity amongst IID cases of all ages. As discussed in Study 2 (Chapter 5), previous studies conducted in the USA and UK have observed inequalities in IID-related hospital admission duration amongst children (Ma, El Khoury and Itzler, 2009; Conway, Phillips and Panday, 1990). However, a Canadian study did not find a statistically significant association between

SES and admission duration for children with rotavirus, when regularly seeing a physician for a medical condition (a proxy measure of comorbidity) was accounted for (Ford-Jones et al., 2000).

Following on from this, in the present study, the prevalence of long-term health problems in a neighbourhood appeared to be an important mediator of the association between neighbourhood deprivation and IID admission duration for adults aged 15–64 years, but not for children or older adults aged 65+ years. Additionally, across all age groups the association between deprivation and emergency hospital admission rates for IID was partly explained by the higher prevalence of long-term health problems in the more deprived neighbourhoods. Long-term health problems seemed to explain more of this association amongst adults aged 15–64 years and adults aged 65+ years, compared to children. Whilst the prevalence of long-term health problems tended to increase with advancing age, deprivation was highly correlated with long-term health problems amongst adults aged 15–64 years, more so than for children or adults aged 65+ years. Similarly, a UK-based study found that inequalities in the prevalence of multimorbidity (defined as two or more chronic conditions) were greatest amongst middle-aged adults compared to children and older adults, with young and middle-aged adults living in the most deprived areas experiencing rates of multimorbidity equivalent to those aged 10–15 years older in the most affluent areas (Barnett et al., 2012). It may be that inequalities in the prevalence of certain chronic diseases level out with advancing age, perhaps due to the dominance of aging as a risk factor, or increased mortality amongst the most disadvantaged at younger ages. This might explain why the prevalence of long-term health problems seemed to explain more of the associations between deprivation and the hospitalisation outcomes amongst adults aged 15–64 years. Detailed information on long-term health problems was unavailable for this study, therefore future research may wish to identify which long-term health problems are most influential in explaining inequalities in hospitalisation outcomes for IID.

Additionally, the associations between neighbourhood deprivation and emergency hospital admission rates for IID were partly explained by geographical factors such as the closer proximity of the more deprived neighbourhoods to hospitals with A&E departments. More deprived neighbourhoods were also more likely to be classified as ‘urban’, and emergency hospital admission rates for IID were statistically significantly higher in urban compared to rural neighbourhoods. Busby, Purdy and Hollingworth (2017b) found admission rates for dehydration and gastroenteritis (definition included specified non-infective gastroenteritis) were higher amongst GPs located closer to A&E departments. Other UK-based studies have also found associations between urban areas, shorter distances from hospital and increased

rates of general (all cause) and respiratory emergency hospital admissions (Bankart et al., 2011; Purdy et al., 2011). These findings might relate to ease of access, since in general the geographical variables appeared to play less of a role in explaining the duration of admission for IID. However, neighbourhoods that were situated closer to GPs were associated with longer admission durations for children and adults aged 65+ years. This finding cannot be easily explained, as it was hypothesised that those living closer to GPs may access primary care services more readily and/or at an earlier stage of IID progression, resulting in shorter admission durations for IID. It is worth noting, however, that people may not choose to be registered with the GP that is situated closest to their place of residence.

A final observation is that age appeared to modify the relationship between ethnicity and the hospitalisation outcomes. Neighbourhoods that had a greater proportion of children of White ethnicity, tended to have higher IID emergency hospital admission rates for children, however the direction of this association was reversed for older adults. Additionally, neighbourhoods that had a greater proportion of children of White ethnicity were associated with statistically significantly shorter admission durations for children, but again the direction of this association was reversed for adults aged 65+ years. In the IID2 Cohort study, rates of IID occurring in the community were not statistically significantly different between those of White ethnicity compared to Mixed, Black, Asian or Chinese ethnicity, however the numbers of participants of Non-White ethnicity were small (Tam et al., 2012b). Without more detailed information on ethnicity, it is difficult to interpret the findings of the present study.

Migration variables were not analysed within this study. Some evidence from the USA and Canada suggests that recently arrived migrants tend to be in better health compared to the non-migrant population, but that over time, the health of migrants tends to deteriorate (Rechel et al. 2013; Domnich et al., 2012). Theories explaining this include a health selection hypothesis, which suggests that migrants tend to be better educated, less risk exposed and thus healthier compared to their compatriots who do not migrate (Domnich et al., 2012). Over time migrants may adopt the norms, values and unhealthy behaviours of the receiving society, or they may find it more difficult to achieve healthy lifestyles due to socioeconomic constraints, which may lead to deteriorations in health outcomes (Jayaweera, 2014). Additional possible explanations include barriers to accessing healthcare, such as inadequate information for recently arrived migrants who may be unfamiliar with healthcare systems in receiving countries (Jayaweera, 2014). It could be said that these theories could potentially contribute towards explaining the findings observed between age, ethnicity and

the hospitalisation outcomes in the present study, however this would be highly speculative given the obvious distinction between ethnicity and migration status.

Other contextual factors which were not investigated (due to data limitations), but could help to explain the socio-spatial gradients in the hospitalisation outcomes for IID, include factors related to the physical environment, community cohesion, and social norms and values. For example, risk of hospitalisation for IID may be associated with the level of social cohesion in a neighbourhood, with close-knit communities providing support and care for those who are unwell, reducing the need for hospitalisation. On the other hand, infections may be transmitted more easily in communities with strong social networks, where individuals visit each other and meet frequently. Additionally, differences between neighbourhoods in the quality of the local food environment may influence admission rates, as food establishments with the lowest hygiene ratings tend to be more concentrated in the most deprived neighbourhoods in England (Collins, 2013).

There are several limitations to this study that should be considered when interpreting the results. Firstly, due to data limitations, admissions were measured as Provider spells instead of CIP spells, and therefore patients who were admitted to one hospital and then transferred to a different hospital may have been counted more than once and admission duration for these patients may have been underestimated. However, a study which compared various methods used to calculate admissions from HES data, found that for dehydration and gastroenteritis (including specified non-infective gastroenteritis) the difference between counts of Provider and CIP spells was minimal (0.3%), and average length of stay was 5.5% shorter when calculated using Provider compared to CIP spells (Busby, Purdy and Hollingworth, 2017a). It was therefore assumed that the use of Provider instead of CIP spells would have had a minimal impact on the results.

Other limitations related to the availability of data, were that the age groups were aggregated to a fairly high level, which may have masked some age specific trends. For example, Olowokure et al. (1999) found that hospital admission rates for IID were over 16 times higher amongst children aged 0–4 years compared to those aged 5–14 years in the West Midlands. Furthermore, information on repeat admissions by the same individual and hospital identifiers were not available within the dataset, which precluded the investigation of clustering at the individual and hospital level. This might have influenced the results since previous studies have found that adults aged 65+ years in particular have a propensity for recurrent IID (Tam et al., 2013). Additionally, following adjustment for clinical and demographic factors including deprivation, Heys, Rajan and Blair (2017) found statistically

significant differences between two London-based hospitals in paediatric length of stay for all causes. These results suggest local variation in the delivery of healthcare between hospitals. Such factors were not taken into account in my analysis, which should be considered when interpreting the results. Having said this, using individual-level data Adams et al. (2017) found that inequalities in the risk of IID persisted even after accounting for recurrent episodes. Similar findings were observed in Study 2 of this thesis when investigating inequalities in IID severity whilst accounting for recurrent episodes. These findings might suggest that if recurrent IID was accounted for, inequalities in the hospitalisation outcomes would remain, however due to data limitations this remains unverified.

In terms of the study design, methodological limitations of ecological studies can include ecological bias whereby associations present at the group-level are not apparent at the individual-level, possibly due to unmeasured confounding or measurement error (Greenland and Robins, 1994). Nonetheless, data were aggregated to relatively small areas (LSOAs containing 1000 to 3000 people) which likely limited the effects of ecological bias. Additionally, because ecological studies are able to capture risk factors and exposures that operate at the community-level (Pearce, 2000), it could be argued that ecological studies are in fact more appropriate for the study of infectious diseases compared to individual-level studies. Individual-level studies may however be best placed to examine some of the associations observed in this study in greater detail, for example the impact of ethnicity and the potential mediating effects of long-term health problems on inequalities in the hospitalisation outcomes for IID.

Lastly, in general the use of routine data for research purposes can have certain drawbacks in that the data are usually collected by numerous individuals who may record data items differently, which can result in measurement error and data quality issues. Particularly pertinent to this study was that a change to the coding of unspecified acute gastroenteritis occurred mid-way though the study period, and uptake of this change may have differed amongst clinicians. However, robustness tests revealed similar results when cases of unspecified acute gastroenteritis were excluded from the analysis. Furthermore, the use of HES data enabled the analysis of all hospital admissions for IID in England. Conducting a similar study of this scale without HES data would not be economically or practically feasible.

In conclusion, this analysis found that increasing neighbourhood income deprivation was statistically significantly associated with increasing emergency hospital admission rates for

IID and increasing duration of admissions for IID, for all ages. Given that most hospital admissions for IID are considered to be preventable, the findings from this study are particularly disconcerting. Unplanned admission to hospital can be distressing and disruptive for patients and families (Pelander and Leino-Kilpi, 2010; Diaz-Caneja et al., 2005), and can incur costs such as loss of work/income which might have more damaging effects for those in lower social positions who have less financial cushioning (Diderichsen, Evans and Whitehead, 2001). Reducing IID hospital admission rates and admission durations experienced by the most disadvantaged to levels experienced by the least, should be an important goal for any levelling-up policy or intervention designed to improve equity in health (Whitehead and Dahlgren, 2007). Further research is warranted to assess the extent to which factors relating to primary care quality might explain the association between deprivation and emergency hospital admission rates for IID.

Chapter 7

Discussion

At the start of this thesis, I presented an initial review of the literature (Chapter 2) which found some evidence of a social gradient in healthcare presentation for GI infections in high income countries. Some studies found those of lower SES compared to high were more likely to present to primary and secondary care services with a GI infection. The three studies presented in this thesis aimed to build upon the literature discussed in Chapter 2, in order to expand current understanding of inequalities in the consequences of GI infections. Significant contributions to the existing knowledge base were made by investigating socioeconomic inequalities in secondary healthcare use for GI infections in greater depth compared to previous studies, and by exploring various possible explanations for the apparent social gradient in healthcare use for IID, such as differential risk of infection or disease severity by SES.

In this final chapter, I provide an overview of my findings in relation to objectives 1–3 of this thesis, and discuss the ways in which the work presented has made a unique and original contribution to the literature. The findings of the studies have been discussed individually within the results chapters, and so this chapter seeks to bring the results of the three studies together, in order to develop greater insight into the social patterning of the consequences of GI infections. The limitations of the research are considered, as well as unanswered questions and future research recommendations. I also address objective 4 of the thesis with a section on the policy implications arising from the studies.

7.1 Overview of findings

In this thesis, I have presented the results of three studies. Each study has addressed the objectives of this thesis set out in Chapter 1. The results of the three studies suggest that socioeconomic inequalities are present in the risk of acquiring an infection amongst children, and in various consequences of GI infections, such as symptom severity, sickness absence and emergency hospitalisation. The key findings from these studies in relation to the objectives are summarised below.

Objective 1 – To systematically review current evidence on the relationship between SES and the incidence of symptomatic GI infections in high income countries, using studies that have identified cases via healthcare records, laboratory notifications and population-based surveys

Study 1 addressed the first objective of this thesis with a systematic literature review and meta-analysis, which included 102 studies that investigated the association between SES and symptomatic GI infection risk in high income countries. The key findings were:

- Age was identified as a statistically significant modifier of the association between SES and the risk of symptomatic GI infections.
- Children (aged <18 years) of lower SES, but not adults, had a greater risk of infection. For children, the risk of GI infection was significantly higher for those of lower SES versus high (RR 1.51; 95% CI 1.26–1.83), but for adults the risk was non-significantly lower (RR 0.79; 95% CI 0.58–1.06).
- For studies that identified cases via hospitals, the pooled risk ratio was 1.47 (95% CI 1.19–1.82), compared to population-based surveys 1.07 (95% CI 0.88–1.29), however the majority of studies that analysed hospitalised cases only analysed children.

Objective 2 – To investigate the association between SES and self-reported IID symptom severity and sickness absence, using data collected in the Second Study of Infectious Intestinal Disease in the Community (IID2 study) in the UK

Study 2 addressed the second objective of this thesis with a cross-sectional analysis of data collected in the population-based UK IID2 study. Data from 1164 individuals with IID aged ≥5 years were analysed. The key findings were:

- Individuals with IID of low SES versus high had twice the odds of experiencing severe symptoms (OR 2.2; 95% CI 1.66–2.87).
- The odds of sickness absence due to IID were also greater for cases of lower SES compared to high (OR 1.8; 95% CI 1.26–2.69).
- The association between SES and sickness absence was attenuated and rendered non-significant after adjusting for symptom severity (OR 1.4; 95% CI 0.92–2.07), indicating that the association between SES and sickness absence was largely explained by differences in symptom severity.

Objective 3 – To assess the impact of neighbourhood income deprivation on emergency hospital admission rates for IID and the duration of these admissions in England, using HES data

Study 3 addressed the third objective of this thesis with a cross-sectional ecological analysis of routinely collected HES data from over 30,000 neighbourhoods in England. The key findings were:

- Increasing neighbourhood income deprivation was statistically significantly associated with increasing emergency hospital admission rates for IID and increasing duration of admissions for IID, for children aged 0–14 years, adults aged 15–64 years and adults aged 65+ years.
- For all age groups, the association between deprivation and admission rates for IID were attenuated but remained statistically significant after controlling for factors such as the higher prevalence of long-term health problems in the more deprived neighbourhoods, and the closer proximity of the more deprived neighbourhoods to hospitals with A&E departments.
- The prevalence of long-term health problems appeared to be an important mediator of the association between deprivation and IID admission duration for adults aged 15–64 years, but not for children or older adults aged 65+ years.

7.2 WHAT THIS THESIS CONTRIBUTES TO THE LITERATURE

In this section, I examine the ways in which the work in this thesis has contributed to our current understanding of inequalities in the risk of GI infections, IID severity and hospitalisations for IID. I also reflect on the novel application of Diderichsen's model of the mechanisms of health inequality (Diderichsen, Evans and Whitehead, 2001) to enhance understanding of inequalities in GI infections. In this discussion, novel perspectives are gleaned by evaluating and synthesising the findings of the three studies presented in this thesis, and considering the findings in the context of other relevant literature. These new perspectives are used to address the overarching aim of this thesis; to enhance current understanding of the extent of socioeconomic inequalities in the consequences of GI infections, and to explore possible explanations for any inequalities identified.

Understanding inequalities in the risk of GI infections

Study 1 was primarily designed to assess the extent of inequalities in the risk of GI infections, but it was also highly informative when interpreting the results of studies examining inequalities in consequences such as healthcare utilisation for GI infections. This is because any social gradient in healthcare presentation for GI infections may reflect increased disease incidence amongst those of lower SES compared to high. Thus, gaining a better understanding of the social patterning in the risk of infection enhances our understanding about inequalities in the consequences following infection.

Study 1 provided a unique contribution to the literature by quantifying, for the first time, the extent of the association between SES and risk of GI infections in high income countries. A number of studies had previously investigated this association, however meta-analytic methods or harvest plots had not been utilised beforehand to assess the relationship. Study 1 expanded on work by Newman et al. (2015) who conducted a systematic review of 16 studies to investigate inequalities in laboratory confirmed foodborne GI infections. Study 1 was designed to identify a far broader set of studies compared to those identified by Newman et al. (2015), including those which identified GI infection cases via population-based surveys, GPs and hospitals, as well as laboratory records. Using this broad approach, it was possible to identify and include 102 studies that measured the association between SES and GI infection risk. This permitted the investigation of the potential modifying effects of several variables on the relationship between SES and GI infection risk, which lead to the generation of truly novel insights, as subsequently discussed.

Modifiers of relationship between SES and symptomatic GI infection risk

A novel analysis was conducted by investigating several potential modifying factors of the relationship between SES and GI infection risk in high income countries. These were pathogen type (based on mode of transmission); country (based on climate and level of development); age of the participants (children aged <18 years, or adults ≥18 years); measure of SES used (area-level or individual-level measures); and the methods used to sample GI infection cases (population-based surveys, laboratory records, GP presentations or hospital admissions). Of these, only age was identified as a statistically significant potential effect modifier of the relationship between SES and GI infection risk in multivariate meta-regression analysis.

It is well recognised that children in general have an increased risk of symptomatic GI infections compared to adults. The UK IID2 Cohort study found children aged <1 year had the highest IID incidence rate of all age groups (1079 cases per 1000 person-years; 95% CI 750–1553), followed by children aged 1–4 years (713 cases per 1000 person-years; 95% CI 603–843) (Tam et al., 2012b). For comparison, the highest incidence rate for adults was observed for those aged 25–34 years, at 335 cases per 1000 person-years (95% CI 268–418) (Tam et al., 2012b). These findings may reflect the fact that children, and infants in particular, have relatively immature immune systems compared to adults (Simon, Hollander and McMichael, 2015), and thus are more vulnerable to infection (Lund and O'Brien, 2011). Furthermore, behaviours of young children, such as crawling and placing objects in their mouths, may increase their exposure to pathogens.

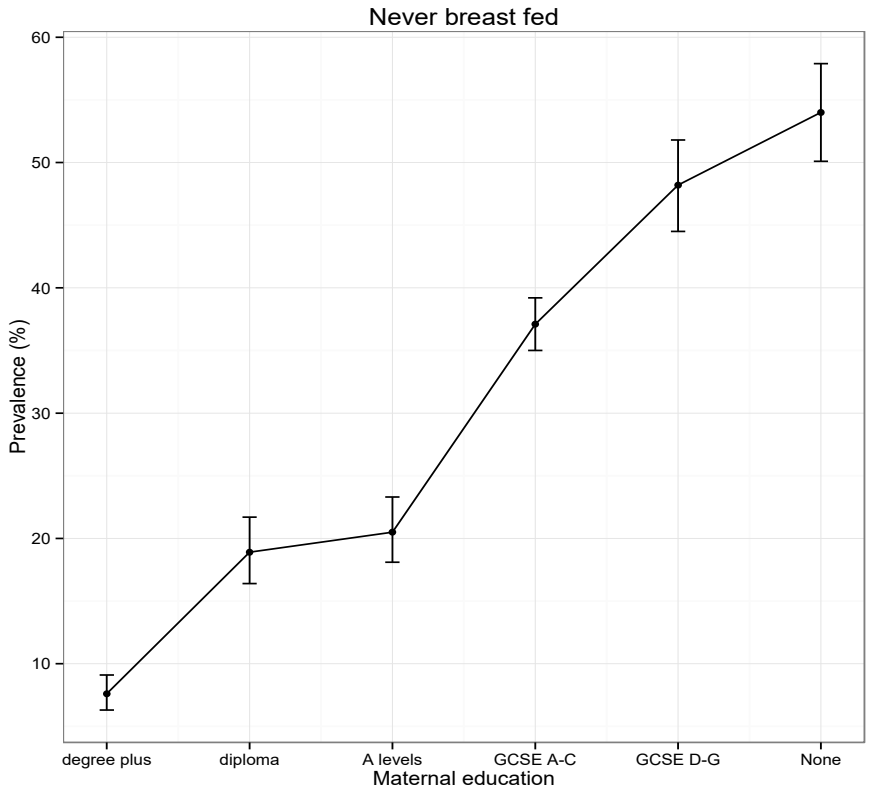
Study 1 in this thesis adds to this knowledge base by suggesting that children of lower SES are particularly vulnerable to GI infections, a finding that was not observed for adults of lower SES. This observation was found by pooling the results of several studies, some of which were population-based surveys, and others were studies that sampled GI infection cases from laboratories, GPs or hospitals. In this respect it can be argued that the results of Study 1 reflect inequalities in the consequences following infection (such as healthcare utilisation) rather than purely assessing inequalities in exposure to and risk of infection. However, by using multivariate meta-regression analysis, it was possible to demonstrate that age significantly modified the relationship between SES and GI infection risk even after controlling for the methods the studies used to sample cases.

The reasons why children of lower SES compared to high were found to be at an increased risk of GI infections in Study 1 might relate to differential exposure to GI pathogens or differential vulnerability/susceptibility to infection by SES. There are several biologically plausible mechanisms by which low SES might influence a child's susceptibility to infection. I consider a number of socially patterned factors that have potential to comprise the immune system in the following section on inequalities in disease severity, however some of these factors are especially relevant here.

Particularly pertinent to the present discussion, is the importance of breastfeeding in infancy. Breastfeeding confers innate and specific immunoprotection against GI infections (Stuebe, 2009). Breast milk contains oligosaccharides that inhibit pathogen binding to host intestinal cells (Newburg, Ruiz-Palacios and Morrow, 2005), and antibodies that provide specific protection against pathogens that are likely to be encountered in an infant's environment (Dieterich et al., 2013). Furthermore, powdered infant formula is not sterile (Lund and

O'Brien, 2011), therefore bottle feeding may increase exposure to pathogens compared to breastfeeding. A number of observational studies conducted in high income countries have found that breastfed children compared to formula fed have a reduced risk of GI infections (Duijts et al., 2010; Chien and Howie, 2001). A UK-based cohort study also found that breastfeeding for 13 weeks or more, protected infants against GI infections throughout the one year follow-up period and beyond the period of breastfeeding itself, controlling for social class, maternal age and parental smoking (Howie et al., 1990). There is also evidence from numerous high income countries indicating that breastfeeding is socially patterned (Oakley et al., 2013; Flacking, Hedberg Nyqvist and Ewald, 2007; Heck et al., 2006; Yang et al., 2004). Analysis of data from the UK-based Millennium Cohort Study (a longitudinal study of over 11,000 infants born in 2001) indicates there is a clear social gradient in breastfeeding in the UK (Figure 7.1) (Taylor-Robinson, 2017). Therefore breastfeeding might be an important mediator of the association between SES and GI infection risk in children, with reduced breastfeeding rates in lower socioeconomic groups increasing infants' susceptibility to GI infections.

Figure 7.1 Social gradient in breastfeeding in the Millennium Cohort Study



Source: Taylor-Robinson (2017)

Additionally, antibiotic use may mediate the association between SES and GI infection risk in children. Evidence from countries such as Scotland, Germany, Sweden and Denmark has shown that antibiotic prescribing tends to be higher amongst children of lower SES (Covvey et al., 2014; Koller et al., 2013; Mangrio et al., 2009; Thrane et al., 2003). Antibiotic use can disrupt the gut microbiome, and increase susceptibility to infections and diarrhoea (Langdon, Crook and Dantas, 2016).

Other factors that might help to explain the findings from Study 1, are inequalities in rotavirus vaccination uptake. Such inequalities might be more pronounced in countries where the rotavirus vaccine is not offered free or fully free of charge. A study conducted in Belgium, where parents are required to provide a partial payment for the vaccine, found children whose parents were unemployed, were less likely to be fully vaccinated against rotavirus (Braeckman et al., 2014). In the UK, the rotavirus vaccine is offered free of charge and coverage in England is around 90%, however work is ongoing as to whether vaccine uptake varies by SES (PHE, 2015; Hungerford et al., 2014). Nonetheless, in Study 1 only one study by Wilking et al. (2012) found evidence of a social gradient in the risk of rotavirus infection amongst children. Therefore whether or not vaccine uptake varies by SES, it seems unlikely that inequalities in the risk of rotavirus infection amongst children could explain the findings observed in Study 1.

In addition to factors that influence susceptibility to GI infections, there are several socially patterned environmental and behavioural factors that might affect children's exposure to GI infections. For example, exposure to pathogens transmitted via the person-to-person route might be increased in situations where households are overcrowded, since overcrowding likely increases the number of contacts that household members have over a period of time (Baker et al., 2013). A systematic review which pooled the results of studies conducted in the USA, UK and South Africa that controlled for age and SES, found that children aged less than five years living in more crowded households experienced an increased risk of gastroenteritis (Baker et al., 2013). Survey and census data from the UK and USA also indicate that household overcrowding is correlated with low SES (ONS, 2014; Solari and Mare, 2012). Evidence from Study 1 provides some support of this theory, where it was found that the risk of GI infection for low compared to high SES was on average higher among studies that analysed pathogens transmitted via the person-to-person route, compared to studies that analysed foodborne pathogens.

Other environmental factors that might mediate the association between low SES and increased exposure to GI infections, include the availability of fast-food from outlets with

poor hygiene ratings in more deprived neighbourhoods. A study by Collins (2013) found that in England, food establishments with the lowest hygiene ratings were more concentrated in the most deprived neighbourhoods. This finding was partly explained by the higher prevalence of takeaway outlets in more deprived areas, since takeaway premises had worse than average food hygiene ratings compared to other food establishments such as restaurants and caterers (Collins, 2013). The impact of this finding on GI infection risk could be sizeable, considering the evidence from one study that found over 50% of children from a deprived UK neighbourhood purchased fast-food twice or more a week (Patterson, Risby and Chan, 2012). Similarly, food hygiene practices inside the home might be socially patterned, however a narrative literature review found a lack of consensus regarding the impact that education or income had on domestic food safety knowledge or risky food handling behaviours in high income countries (Al-Sakkaf, 2015).

Apart from inequalities in breastfeeding and vaccination uptake, the potential explanations for the observed inequalities in GI infection risk in children that have been discussed, could equally be applied to adults. However, findings from Study 1 suggest that adults of low and high SES have a similar risk of symptomatic GI infections. As suggested in Chapter 4, this result could reflect differential immunity by SES. Antibodies that are produced in response to a GI infection, can provide protection against future disease for several years following the initial infection (Simmons et al., 2013; Miller et al., 2005). If children and adults of lower SES are more frequently exposed to enteric pathogens, they may be more likely to develop an acquired immunity to subsequent disease compared to those of higher SES. Studies from the USA and Sweden indicate that amongst poultry abattoir workers, the majority of campylobacteriosis cases occur amongst new employees, and more experienced employees have higher levels of circulating antibodies (de Perio et al., 2013; Cawthraw et al., 2000; Christenson et al., 1983). These results suggest that long-term exposure to certain pathogens might provide some level of immunity to future disease (Janssen et al., 2008). Thus, children of lower SES might be more exposed or susceptible to GI infections, but increased exposure to enteric pathogens in childhood and adulthood may increase immunity and resistance to subsequent IID in individuals of lower SES.

Understanding inequalities in IID severity

The second and third studies in this thesis generated novel insights by analysing inequalities in IID severity for all age groups, and investigating the potential mediating role of IID

severity on the association between SES and sickness absence due to IID. The results of these studies and their original contribution to the literature are discussed below.

Inequalities in sickness absence due to IID

The results of Study 2 showed that in the UK individuals with IID of lower SES compared to high tend to report more severe IID symptoms and sickness absence, and symptom severity appears to be an important mediator of the association between SES and absence. The finding that the association between SES and sickness absence was largely explained by greater symptom severity amongst more disadvantaged groups, is a new contribution to the literature. Only two studies could be identified that have previously investigated the relationship between SES and sickness absence due to IID in high income countries (Mohren et al., 2005; Feeney et al., 1998). These studies, conducted in Britain and the Netherlands produced inconsistent results. The British study by Feeney et al. (1998) found a social gradient in age adjusted rates of sickness absence due to gastroenteritis amongst UK civil servants, whereas the Dutch study by Mohren et al. (2005) found no such association in univariate analysis. However, neither study examined the potential mediating role of IID severity on the relationship between SES and sickness absence, which was explored in Study 2 of this thesis.

Studies conducted in a number of high income countries have found that rates of general (all cause) sickness absence amongst employees are higher for those of lower SES compared to high (Kristensen et al., 2010). However, some of these studies have demonstrated that this association can in part be explained by the increased levels of morbidity for those of lower SES (Kaikkonen et al., 2015; Johansson and Lundberg, 2009). Other factors that have been found to partly explain the social gradient in all cause sickness absence include physical working conditions (such as hazards in the workplace, increased levels of noise and physically demanding work), and psychosocial working conditions (such as job insecurity, low support at work, high job demands and low control) (Niedhammer et al., 2017; Kaikkonen et al., 2015; North et al., 1993).

Loss of working days due to sickness or caring for an unwell child, can have important economic consequences and these are likely to be more severe for more disadvantaged groups who might receive less adequate compensation from their employer. For example, self-employment is rising in the UK, however individuals who are self-employed are not covered by employment law meaning they are not entitled to receive the National Minimum Wage or Statutory Sick Pay (Field et al., 2017). As has been recently highlighted in the

media, companies can even charge self-employed workers if they are off sick and cannot find replacement cover (Davies, 2017). Thus sickness absence can incur loss of earnings and financial penalties. These consequences have the potential to widen income inequalities given that there appears to be a social gradient in sickness absence due to IID.

Additionally, loss of days from school can affect educational attainment. A cohort study conducted by the Department for Education, that analysed data from pupils attending state-funded schools found a weak but statistically significant effect of school absence on educational attainment at Key Stages 2 and 4, controlling for prior attainment, ethnicity, sex, special educational need and free school meal eligibility (Department for Education, 2016). For each half day absent, reductions of around 0.4% and 1.8% in the odds of achieving Level-5 or above in reading and mathematics at Key Stage 2, and achieving five A*–C grades at General Certificate of Secondary Education (GCSE) level were observed, respectively (Department for Education, 2016). Receiving free school meals (a proxy measure of low SES) was associated with reductions of 20.7% and 43% in the odds of achieving Level-5 or above at Key Stage 2, and five A*–C GCSEs, respectively. These results suggest that the unequal effects of IID on sickness absence from school could exacerbate educational inequalities in childhood, which can influence subsequent socioeconomic position in adulthood.

Inequalities in IID severity

The results of Study 2 also showed that in the UK individuals with IID of lower SES compared to high tend to report more severe symptoms. These findings were observed for children and adults. At the more severe end of the disease spectrum, Study 3 demonstrated a socio-spatial gradient in the duration of hospital admissions for IID in England, and again this finding was observed across all age groups. These findings taken together suggest that in the UK, whether or not IID is relatively mild or relatively severe requiring hospitalisation, those of lower SES compared to high tend to experience greater levels of disease severity in terms of the symptoms they experience and the duration of these symptoms.

In terms of inequalities in disease severity, the previous literature that was identified on this topic in high income countries, had exclusively studied paediatric cases, usually in hospital settings (Ma, El Khoury and Itzler, 2009; Ford-Jones et al., 2000; Baker, Taylor and Henderson, 1998; Conway, Phillips and Panday, 1990). The findings from this thesis generated original insights by analysing inequalities in IID severity for all age groups, and confirmed those of previous studies conducted in paediatric populations. Studies by Ma, El

Khoury and Itzler (2009) and Conway, Phillips and Panday (1990) found social gradients in the duration of hospital admissions for American and British children, respectively. Interestingly, studies by Ford-Jones et al. (2000) and Baker, Taylor and Henderson (1998) found non-statistically significant associations between SES and the duration of symptoms amongst children, when they controlled for potential mediators of the association (the presence of comorbidities and breastfeeding, respectively).

In Study 3, the prevalence of long-term health problems in a neighbourhood appeared to be an important mediator of the association between neighbourhood deprivation and IID admission duration for adults aged 15–64 years, but not for children or older adults aged 65+ years. This may have been because deprivation was highly correlated with the prevalence of long-term health problems amongst adults aged 15–64 years, more so than for children or adults aged 65+ years. A number of chronic diseases can cause secondary immunodeficiency such as diabetes mellitus, chronic kidney disease, HIV and chronic lymphocytic leukemia (Kurts et al., 2013; Chinen and Shearer, 2010; Goldman, 2000). Additionally, management of cancers and several autoimmune inflammatory conditions, such as rheumatoid arthritis and inflammatory bowel disease, often involves the use of immune suppressing treatments and medications (Chinen and Shearer, 2010; Hsu and Katelaris, 2009). These factors may increase susceptibility to and severity of IID, however it is somewhat difficult to draw inferences from the results of Study 3 without more detailed information on the long-term health problems captured.

In addition to comorbidity, there are several other biologically plausible mechanisms by which low SES might influence IID severity. Socially patterned factors that have potential to compromise the immune system, such as smoking, increased levels of chronic stress and nutritional deficiencies, might mediate the association between SES and IID severity. As discussed in the previous section, these factors might also influence a person's vulnerability/susceptibility to a GI infection, as well as the severity of disease they experience following infection (Lund and O'Brien, 2011).

For instance, socioeconomic inequalities in smoking rates are well established, and cigarette smoke is known to have immune suppressing properties. Research suggests that whilst smoking uptake has reduced in the UK since the 1990s, inequalities in smoking initiation amongst adolescents remain; a finding that is largely explained by regular exposure to an adult smoker in the same room (Taylor-Robinson et al., 2017; Green et al., 2016). Additionally, research has shown that tar and nicotine can act as immunosuppressants; increasing a host's susceptibility to infections by modifying innate immune responses

(Mehta, Nazzal and Sadikot, 2008; Sopori, 2002). These immune suppressing effects can occur in active smokers as well as passive smokers (Stämpfli and Anderson, 2009), such as children exposed to smoke in the home environment. These findings suggest that for both children and adults, low SES might influence IID severity via the mediating effects of smoking.

It is also conceivable that psychosocial factors such as chronic stress might mediate the association between SES and IID severity. Research suggests that individuals of lower SES compared to high are more routinely exposed to psychosocial stressors in their living and working environment, contributing to a persistent level of background stress (Lantz et al., 2005; Steptoe and Feldman, 2001; Baum, Garofalo and Yali, 1999). Chronic stress in childhood can disrupt the self-regulatory processes that help children cope with external demands, and as such the effects of chronic stress in childhood can have lasting consequences into adulthood (Evans and Kim, 2013; Evans and Schamberg, 2009). Chronic stress has been shown to have negative effects on immune system functioning, suppressing the body's ability to initiate an efficient immune response to infection (Salleh, 2008; Segerstrom and Miller, 2004).

Furthermore, the effects of material deprivation may influence factors such as nutrition. Evidence from the UK and other high income countries suggests fruit and vegetable consumption is lowest amongst lower socioeconomic groups, and intakes of most vitamins and minerals are higher in more affluent groups (Miller, Spiro and Stanner, 2016; Maguire and Monsivais, 2015; Darmon and Drewnowski, 2008). Results of a fairly recent literature review suggest these findings may relate to affordability, since nutrient-dense foods are generally more expensive per calorie than energy-dense foods (Darmon and Drewnowski, 2015). Deficiency of some micronutrients, even when relatively mild, can alter immune responses (Chandra, 1997). For example, zinc, iron, copper, selenium and vitamins A, C and E are required for efficient immune system functioning, and inadequate intake may lead to suppressed immunity which increases susceptibility to infections (Wintergerst, Maggini and Hornig, 2007; Chandra, 1997). In this way, those of lower SES compared to high may be at increased risk of severe IID, due to inadequate diet.

In addition to these biologically plausible factors, from an educational perspective the awareness of the importance of re-hydration may differ across socioeconomic groups. Individuals of low SES may lack knowledge about the appropriate treatment for IID, which might put them at a greater risk of dehydration. This theory is however speculative and further research is needed in this area.

Understanding inequalities in healthcare utilisation due to IID

Study 3 in this thesis also examined inequalities in emergency hospital admission rates due to IID in England. This study was designed to address a gap in the literature identified in Study 1, where it was found that the majority of studies that had assessed inequalities in hospital admissions for IID in high income countries, had also focused on paediatric populations. In the UK, two ecological studies had previously investigated hospitalisations for IID in children, both of which aggregated data over large areas (PCT-level) (Kyle et al., 2011; Pockett et al., 2011). Additionally, one study had examined IID-related hospital admissions for five age groups ranging from ages 0–4 years to 75+ years, however the data analysed were collected 20 years ago and were limited to the West Midlands region in England (Olowokure et al., 1999).

Study 3 made a unique contribution to the literature by providing an up-to-date measure of the association between neighbourhood deprivation and emergency hospital admission rates for IID across England, for three age groups (children aged 0–14 years, adults aged 15–64 years and adults aged 65+ years), using small area measures to aggregate data. Using data linkage techniques, it was also possible to investigate the effects of several neighbourhood-level characteristics on the relationship between deprivation and emergency hospital admission rates for IID. The results of Study 3, confirmed those of Pockett et al. (2011) and Olowokure et al. (1999) observed in the UK previously. A socio-spatial gradient in emergency hospital admission rates for IID was observed for all age groups, over the seven year period from 2009–15. Additionally, it was found that the associations between deprivation and emergency hospital admission rates for IID were partly explained by factors such as the higher prevalence of long-term health problems in the more deprived neighbourhoods, and the closer proximity of the more deprived neighbourhoods to hospitals with A&E departments. These findings had not been described previously in the UK.

As mentioned, there are several common chronic conditions that can result in secondary immunodeficiency and as such these conditions may increase susceptibility to and severity of IID. Since hospitalisation is in itself a severe consequence of IID, it seems biologically plausible that the higher prevalence of long-term health problems in the more deprived neighbourhoods would mediate the association between deprivation and admission rates for IID. Additionally, more deprived neighbourhoods were more likely to be classified as ‘urban’ and were located closer to hospitals with A&E departments, and these factors explained some of the association between deprivation and emergency hospital admission rates for IID. Previous studies have found similar relationships between closer residential

proximity to A&E hospitals and higher emergency admission rates in general, but less is known about the extent to which geographical/proximity factors explain associations between deprivation and emergency hospital admission rates. It was thought that the findings observed in Study 3 might relate to ease of access, since in general the geographical variables appeared to play less of a role in explaining the duration of admissions for IID. However, it was also found that neighbourhoods situated closer to GPs were associated with longer admission durations for children and adults aged 65+ years; a finding that was not expected and which suggests closer proximity to health services might not solely reflect ease of access. Other studies have expressed uncertainty about whether residents who live in closer proximity to A&E services have higher emergency hospital admission rates in general because they find it easier to access these services or because of wider factors relating to care quality in urban versus rural areas (Purdy, 2010). Further research on this topic may improve understanding of the relationship between geographical factors, ease of access to health services and inequalities in emergency hospital admissions for IID.

Since long-term health problems and geographical factors only partially explained the associations between deprivation and hospital admission rates for IID, the rest of this section is devoted to exploring various additional explanations for the inequalities observed in Study 3, and inequalities in IID-related healthcare use in general. Particular reference is made to the results of the studies presented in this thesis. By synthesising the findings of the three studies from this thesis, and considering the evidence in the context of other relevant literature, it is intended that novel perspectives will be gleaned to enhance current understanding of inequalities in the consequences of IID.

Risk of infection as an explanation for inequalities in hospitalisation for IID

As mentioned, it was intended that the results of Study 1 would help to enhance understanding of inequalities in consequences such as healthcare presentation for IID. The findings from Study 1 can be applied to gain a greater comprehension of the results from the third study of this thesis, which found a socio-spatial gradient in hospital admission rates for IID in England.

A recently published literature review found evidence to suggest that in general, those of lower SES compared to high tend to use more healthcare at any given age, because they are sicker (Cookson et al., 2016). If individuals of lower SES have a greater risk of symptomatic GI infections, it stands to reason that they would also have a greater need for healthcare services and thus be more likely to present. However, the results of Study 1 suggest that

differential risk of infection across socioeconomic groups is unlikely to be the sole explanation for the apparent social gradient in hospital admission rates for IID. Evidence from Study 1 suggests that for adults in high income countries there is no statistically significant relationship between SES and risk of infection. On the other hand, children of lower SES were found to have an increased risk of infection. Nonetheless, this finding does not explain the socio-spatial gradient in hospital admission rates observed across all age groups in Study 3. Thus, it would appear that an alternative explanation underlies the socio-spatial gradient in admission rates across all ages. Interestingly, Study 3 also observed a steeper socio-spatial gradient in hospital admission rates for children compared to adults in the fully adjusted models. Thus, it might well be that differential risk of infection explains some of the socio-spatial gradient in hospital admission rates in children, but that an alternative explanation underlies the socio-spatial gradient in admissions across all ages. Certainly for adults an alternative explanation seems more likely.

These inferences are of course hypothetical; no formal tests have been performed to assess whether inequality in hospital admission rates for IID can be explained by inequalities in the risk of infection. Additionally, Study 1 assessed inequalities in the risk of GI infection by pooling the results of studies conducted in a number of high income countries, rather than studies specific to the UK. However, a recent study that utilised longitudinal methods to analyse data from the UK IID2 Cohort study, found that individuals of lower SES compared to high had a statistically significant lower risk of IID (Adams et al., 2017). The results of this study likely give a good indication of inequalities in the risk of infection occurring in the community, since the IID2 study is the largest and most up-to-date population-based survey of IID conducted in the UK. This UK specific study, gives further support to the hypothesis that the social gradient in emergency hospital admissions for IID cannot be entirely explained by inequalities in the risk of infection.

Inequalities in the quality of primary care

Another explanation for the inequalities observed in emergency hospital admission rates for IID might relate to inadequacies in the provision or quality of primary care in deprived neighbourhoods. Several UK-based studies have found that those of lower SES compared to high have higher rates of emergency hospital admissions for conditions where hospital admission is considered to be preventable with effective management at the primary care level (Barker, Steventon and Deeny, 2017; Busby, Purdy and Hollingworth, 2017b; Cookson et al., 2016; Tian, Dixon and Gao, 2012). Such conditions (called ACSCs) include chronic conditions such as asthma, chronic obstructive pulmonary disease and diabetes

complications, as well as acute conditions such as gastroenteritis. Whilst it could be said that inequalities in emergency hospital admission rates for some of these conditions might reflect increased need amongst those of lower SES, the findings from Study 1 of this thesis suggest that the social gradient in emergency hospital admissions for IID cannot be entirely explained by inequalities in the risk of infection. Alternatively, inequalities in the management of IID at the primary care level might help to explain the social gradient in emergency hospital admissions for IID.

Previous studies have shown that improvements to primary care quality can reduce emergency hospital admissions for ACSCs. The Quality and Outcomes Framework (QOF) introduced in 2004 in the UK, is a pay-for-performance scheme that financially rewards GPs for meeting certain care quality targets (Roland and Guthrie, 2016). A longitudinal study conducted in the UK found that the introduction of the QOF was associated with sustained reductions in emergency hospital admission rates for ACSCs that were incentivised under the QOF, however admission rates for ACSCs that were not incentivised (such as dehydration) continued to rise over the study period (Harrison et al., 2014). This suggests that interventions to improve primary care quality and subsequent reductions in hospitalisation outcomes need to be disease/condition specific.

Findings in support of this were observed by a recent UK-based study that analysed predictive factors of emergency hospital admission for 28 different ACSCs. Busby, Purdy and Hollingworth (2017b) found that the impact of GP-level factors (e.g. care quality [measured using QOF indicators], continuity of care, and access to care) on admission rates varied across the ACSCs. Lower continuity of care was associated with higher rates of unplanned hospital admissions for dehydration and gastroenteritis (a definition that included specified non-infective gastroenteritis, e.g. allergic and dietetic gastroenteritis) (Busby, Purdy and Hollingworth, 2017b). In this study continuity of care was measured as relationship continuity with a clinician, and as the results suggest, improvements in this area may lead to reductions in emergency hospital admissions for gastroenteritis. However, questions remain as to whether improving relationship continuity at the primary care level would lead to reductions in *inequalities* in emergency hospital admissions for IID.

It has long been recognised in the UK that the availability of good quality healthcare tends to vary inversely with need (Hart, 1971), however measuring inequalities in primary care quality is complex. A recently published study found that between 2004–5 and 2011–12 inequalities in access to primary care and primary care quality (assessed using QOF indicators) had substantially reduced in England, however large inequalities remained in

healthcare outcomes such as hospitalisations for chronic ACSCs and amenable mortality (Asaria et al., 2016). Another recent UK-based study also reported that improvements to primary care supply and quality were associated with reductions in average all cause emergency hospital admission rates but were not associated with reducing inequality gradients in emergency admission rates (Sheringham et al., 2017). Future research on this topic could try to enhance understanding by assessing the extent to which factors relating to primary care quality might explain the association between deprivation and emergency hospital admission rates for IID.

Pro-poor bias by healthcare professionals in referral and admission decisions

An alternative explanation for the social gradient in hospital admission rates for IID, might be that the care provided by clinicians differs depending on the SES of the presenting patient. The NICE guidelines for clinicians treating children under five years of age with gastroenteritis, recommend repeat face-to-face assessment or referral to secondary care for children whose social circumstances require continued involvement of healthcare professionals (NICE, 2009). General practitioners may therefore be more likely to refer children of lower SES to secondary care services, reflecting a 'pro-poor' bias in referral practices. However, this theory does not explain the socio-spatial gradient in emergency hospital admission rates for IID observed across all age groups in Study 3. Qualitative research is currently being conducted by colleagues from the University of Liverpool to gain a better understanding of healthcare professionals' beliefs and attitudes regarding the management of patients with GI infections in socioeconomically contrasting areas (McGarrol, 2017).

Inequalities in disease severity

The explanations for the apparent social gradient in hospital admissions for IID discussed thus far appear to have certain limitations, and certainly based on the results of Study 1 it seems unlikely that inequalities in the risk of infection can fully explain the inequalities in admission rates observed. Inequalities in disease severity on the other hand might provide a plausible alternative explanation. If individuals of lower SES tend to experience greater levels of IID severity, this might explain some of the social gradient in healthcare utilisation for IID. Results from Studies 2 and 3 of this thesis suggest that whether or not IID is relatively mild or relatively severe requiring hospitalisation, those of lower SES compared to high tend to experience greater levels of disease severity in terms of the symptoms they experience and the duration of these symptoms. Furthermore, the results of these studies

suggest that both adults and children experience inequalities in IID severity. It therefore seems highly plausible that differential IID severity by SES might help to explain the inequalities in primary and secondary care presentation for IID that have been observed in the literature.

Further support for this theory is provided by several studies that have analysed predictive factors for GP presentation amongst IID cases in high income countries. As discussed in Chapter 2, these studies have universally found that IID cases with more severe symptoms are more likely to present to their GP for their illness (Doorduyn, Van Pelt and Havelaar, 2012; Van Cauteren et al., 2012; Adlam et al., 2011; Scallan et al., 2006; Tam, Rodrigues and O'Brien, 2003; De Wit et al., 2001b). Placing the results from this thesis within the context of these studies, provides further evidence that disease severity might mediate the pathway between SES and healthcare presentation for IID. Thus, individuals of lower SES might present to healthcare services more frequently for IID because they indeed have a greater need for such services.

Differential healthcare-seeking behaviours by SES

The possibility that differential healthcare-seeking behaviour by SES might help to explain the social gradient in hospital admissions for IID, has not been explored by the studies presented in this thesis. A UK study by Adamson et al. (2003) who surveyed 1350 individuals, found that those of lower SES compared to high were as likely, if not more likely to report they would access healthcare immediately in response to imaginary case scenarios such as experiencing chest pain and discovering a lump in the armpit. Conversely, a qualitative study of individuals with chest pain conducted in Glasgow found those from deprived areas did not tend to report presenting to their GP with chest pain more frequently compared to individuals from more affluent areas (Richards, Reid and Watt, 2002). Some perceived they were to blame for their condition and would be chastised by their doctor. Whilst these studies do not examine IID specifically, they show how differential healthcare-seeking behaviours by SES might influence healthcare presentation rates.

It could be said that the social patterning of healthcare-seeking behaviours for IID has been examined to some extent by studies that have investigated predictive factors for GP presentation amongst IID cases. For example, in their analysis of the English IID1 study Tam, Rodrigues and O'Brien (2003) found that cases with IID who left full-time education before 16 years of age, were statistically significantly more likely to present to their GP for their illness, compared to cases who left education at 19 years of age or older. These findings

were observed even after adjusting for disease severity. This might indicate that those of lower SES compared to high have an increased tendency to seek healthcare advice for IID, even after accounting for increased illness severity amongst those of lower SES. However, whether this inference can be generalised to secondary care healthcare-seeking is unclear. Certainly, it seems less likely that differential healthcare-seeking by SES could explain the inequalities in emergency hospital admissions for IID observed, since the decision to admit a patient to hospital is made by the attending clinician. This topic could potentially be investigated further by comparing the extent of inequalities in A&E presentation rates with inequalities in admission rates for IID. Alternatively, the role of healthcare-seeking behaviours in terms of explaining inequalities in both primary and secondary care presentation for IID could be investigated using qualitative methodologies to examine decision thresholds for seeking medical advice, and to explore whether these thresholds differ by SES. Current research being conducted by colleagues at the University of Liverpool aims to enhance our understanding of these topics using ethnographic methodologies (Rotheram, 2017).

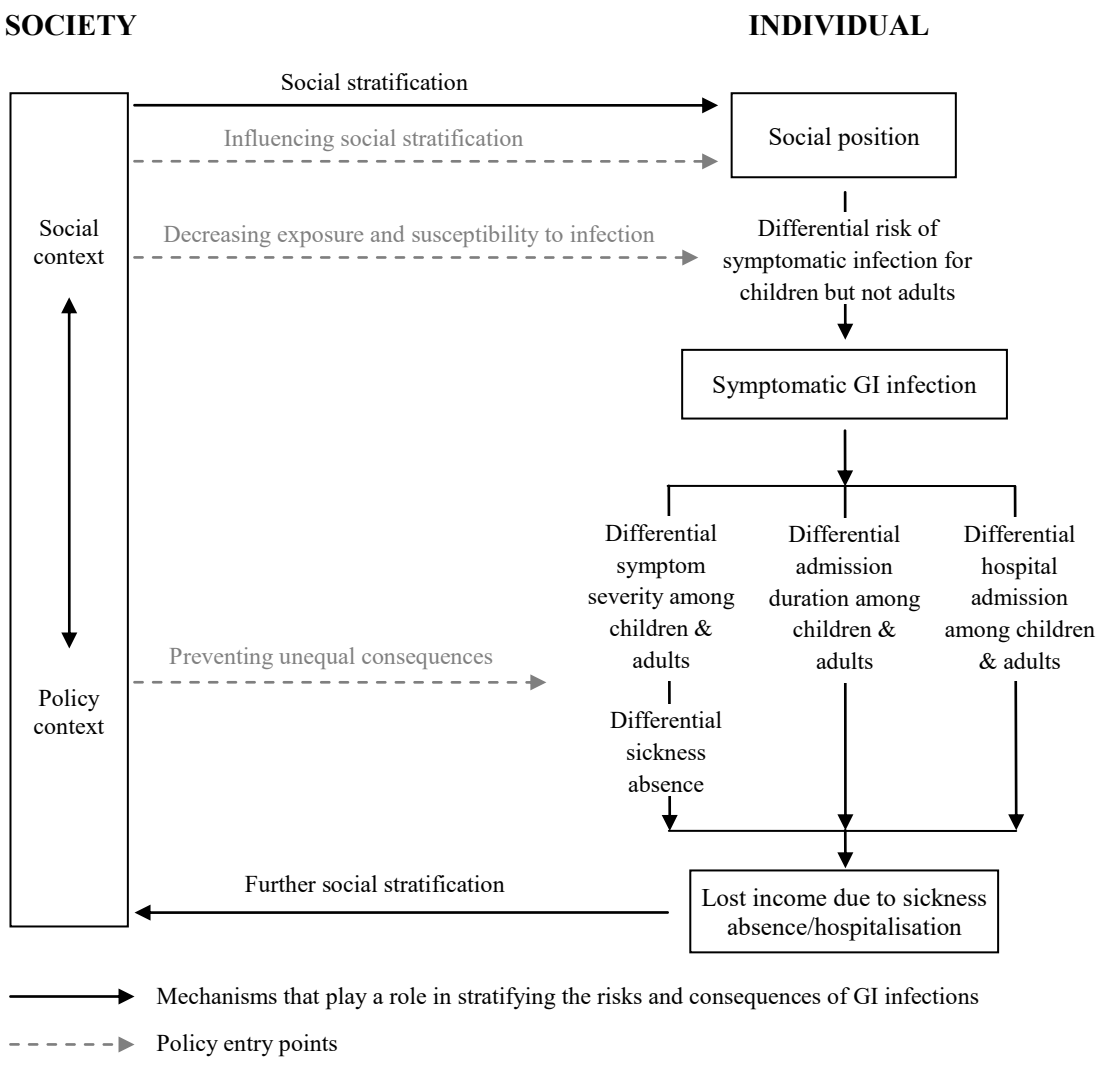
Applying Diderichsen's model to GI infections

Diderichsen's model of the mechanisms of health inequality (Diderichsen, Evans and Whitehead, 2001) has served as a useful theoretical framework on which to base this inquiry. Applying Diderichsen's model to enhance understanding of inequalities in GI infections, is in itself a novel approach. As discussed in Chapter 2, the model outlines the potential mechanisms by which social contexts and positions might influence health outcomes and lead to health inequalities. It demonstrates how a person's social position may determine their exposure to health damaging risks, their vulnerability to ill health following such exposure, and the consequences they experience following a disease event. The model also shows how differential consequences of ill health by SES may result in further social stratification. For example, the impact of lost income due to ill health may have more damaging effects for those in lower social positions who have less financial cushioning, compared to those in higher positions (Diderichsen, Evans and Whitehead, 2001).

The studies presented in this thesis have investigated the mechanisms of health inequality proposed in Diderichsen's model, in the context of GI infections. An adapted model is displayed in Figure 7.2, outlining the findings of this thesis. Study 1 found that social position appears to determine children's exposure and vulnerability to IID, but for adults there appears to be no social gradient in the risk of symptomatic infections. Additionally,

Studies 2 and 3 found evidence to suggest that social position determines a number of consequences of GI infections for both children and adults, such as disease severity, sickness absence, emergency hospital admission rates and the duration of hospital admissions. It seems plausible that some of these consequences, such as sickness absence and hospital admission, may incur financial penalties such as lost income, which may have more damaging effects for those in lower social positions and generate deeper social divides.

Figure 7.2 Diderichsen's model applied to GI infections



Source: Adapted from Diderichsen, Evans and Whitehead (2001)

Thus, important insights that enhance understanding of inequalities in GI infections have been gleaned. The work in this thesis suggests that inequalities in GI infections occur throughout the disease pathway from exposure to and risk of infection, to the consequences

following infection. However, it appears that social position may play a greater role in determining the consequences that occur following infection compared to the risk of infection.

7.3 STRENGTHS AND LIMITATIONS

The strengths and limitations of each study have been discussed in the individual results chapters. In this section I will provide an overview of the key strengths and weaknesses of the studies, and I will highlight in more general terms the limitations of the datasets and methods used within this thesis.

Overall, the work presented in this thesis has made a unique and significant contribution to our current understanding of inequalities in the risks and consequences of GI infections. Examining inequalities in GI infections is of public health importance due to the high frequency with which these infections occur in the community, and the associated burdens and costs for individuals, the healthcare sector and wider economy. Original contributions to the existing knowledge base have been made by investigating the association between SES and GI infection risk in high income countries using meta-analytic methods, analysing inequalities in IID severity for all age groups in the UK, exploring the potential mediating role of IID severity on the association between SES and sickness absence due to IID, and evaluating inequalities in IID-related hospital admissions in England in greater depth compared to previous studies. The data sources used in this thesis are generally of high quality and rigorous methods have been utilised throughout. The main limitations relate to data availability issues which are discussed in more detail below.

Study 1 – Systematic review

Study 1 was a systematic review which aimed to investigate inequalities in the risk of GI infections in high income countries. Key strengths of this study were the use of systematic methods throughout, which increased the reliability and validity of the conclusions drawn. An explicit research question was developed, and clear inclusion/exclusion criteria were adhered to throughout the screening process. The search strategy was extensive, involving electronic and hand searching to identify, as far as possible, all relevant literature. Included studies were individually assessed for quality, and to increase reliability two methods were

used to synthesise the study findings. The use of harvest plots enabled the results of all of the studies to be captured and synthesised, whereas the meta-analytic methods permitted a precise assessment of potential sources of statistical heterogeneity amongst the studies. Several sensitivity analyses were performed to test the robustness of the results, and the screening and quality assessment processes were performed independently by two reviewers which added rigor to the review.

The broad nature of the review was both a strength and limitation of the study. The studies that were included used various measures of SES, definitions of GI infections, analysed different populations and identified cases from different sources. This variation meant that it was possible to investigate the modifying effects of several variables on the relationship between SES and GI infection risk, which lead to the generation of novel insights. However, despite investigating several potential effect modifiers, a considerable amount of statistical heterogeneity in the effect estimates remained unexplained. It is possible that factors that were difficult to adjust for may have explained the high residual heterogeneity. For example, the primary aims of the individual studies varied, as did the variables used to statistically adjust the associations between SES and GI infection risk. Additionally, the categorisation of low and high SES may have differed considerably between studies. The large amount of statistical heterogeneity observed may have also negatively affected the power to detect statistically significant modifiers in the meta-regression (Hempel et al., 2013). Nonetheless, age was identified as a statistically significant effect modifier of the relationship between SES and GI infection risk, and this finding was confirmed by several sensitivity analyses.

In general, by combining the results of several studies on a particular topic, systematic reviews and meta-analyses can establish the generalisability and consistency of an effect (Greenhalgh, 2010). Systematic reviews can therefore provide strong evidence on which to base healthcare/policy decisions and enhance theoretical growth (Ross, 2012; Polgar and Thomas, 2008). However, systematic reviews do not mitigate the effects of bias in the original studies that are being combined, and as such the conclusions of a systematic review are only as reliable as the methods used in these original studies (Garg, Hackam and Tonelli, 2008). This can be a particular issue when combining effect estimates of observational studies compared to randomised-controlled trials, since observational studies can be prone to the effects of bias and confounding (Egger, Smith and Schneider, 2001). Indeed, it has been suggested that the purpose of meta-analysis of observational studies is to explore potential reasons for heterogeneous risk estimates across studies, rather than to obtain an overall summary statistic for the combined studies (Kheifets et al., 1995). This approach was taken for the systematic review presented in this thesis, since all of the studies identified for

inclusion were observational in nature. Additionally, sensitivity analyses were conducted excluding studies classified as being of low quality, which confirmed the results from the main analysis for the age subgroups.

Study 2 – Cross-sectional analysis of IID2 Study dataset

Study 2 was a cross-sectional analysis of data collected in the population-based UK IID2 study, which aimed to investigate inequalities in self-reported IID symptom severity and sickness absence due to IID. A key strength of this study was the use of the IID2 study dataset, which is the largest and most recent population-based survey of IID conducted in the UK. In terms of analysing inequalities in the consequences of IID, identifying IID cases for research purposes via population-based surveys can offer advantages over other methods such as laboratory records, since laboratory notified cases represent only a small fraction of cases occurring in the community (Tam et al., 2012b). Whilst the definition of IID used in the IID2 study was based on the symptoms experienced by participants, it could be argued that when estimating the burden of IID on society and inequalities in this burden, capturing definite and possible cases of IID is preferable to capturing only cases with a laboratory confirmed infection.

Additional strengths included the hierarchical approach used for the multivariate regression modeling. Prior to modeling the data, logic models were created to theoretically demonstrate the way in which the independent variables might be causally related to the outcomes under investigation. These logic models informed the order of entry of the variables into the multivariate regression models. Hierarchical approaches to multivariate modeling require the researcher to carefully plan the analysis based on theoretical knowledge (Malek, Berger and Coburn, 2007). As such, the models that are created can be more informative and have more theoretical and practical value compared to models created using stepwise approaches whereby independent variables are entered into the model based on mathematical criteria (Field, Miles and Field, 2012; Malek, Berger and Coburn, 2007).

Furthermore, several sensitivity analyses were performed to check the robustness of the results. Similar results to those reported in the main analysis were observed when analyses were conducted with recurrent episodes of IID included with clustering at the individual level accounted for using mixed-effects models, and when multiply imputed datasets were analysed. These findings strengthened the results of the study, particularly those from the

multiply imputed datasets since there was a large amount of missing data within the main exposures and outcomes of interest.

Whilst the use of the IID2 study dataset had certain benefits as mentioned, there were certain limitations of the IID2 study which may have reduced the generalisability of the findings of Study 2. Within the IID2 Cohort study individuals in managerial/professional occupations, those aged 55+ years and those of White ethnicity were over-represented compared to the UK population, and individuals in intermediate and routine/manual occupations and those aged 15–24 years in particular were under-represented (Tam et al., 2012b). Under-representation of lower socioeconomic groups is commonplace in population-based surveys (Lorant et al., 2007), and this could limit the external validity of the findings of Study 2.

A limitation of cross-sectional studies in general is that because data on the exposures and outcomes are collected at the same time, it can be difficult to determine the temporal sequence of events (Carlson and Morrison, 2009). Whilst it has been proposed in this discussion that inequalities in certain consequences of IID such as sickness absence and hospitalisation may contribute to greater social stratification, it seems highly unlikely that the consequences of IID are the main determinants of a person's socioeconomic position. When investigating inequalities in acute infections such as IID, researchers can be fairly confident about the temporal sequence of events.

Additionally, whilst every effort has been taken to limit the effects of confounding on the study results, the possibility of unobserved confounding explaining the results of Study 2 cannot be completely ruled out. Nevertheless, the results of Study 2 have been extensively evaluated within the context of previous literature, and the findings are broadly consistent with those of previous studies and plausible explanations for inequalities in IID severity have been identified. This adds weight to the theory that the observed association between SES and IID severity is causal in nature, rather than the effect of bias or confounding (Hill, 1965).

Study 3 – Ecological analysis of HES data

Study 3 was a cross-sectional ecological analysis of HES data which aimed to assess inequalities in emergency hospital admission rates for IID, and in the duration of these admissions. A key strength of this study was the scope and timeframe of the HES data extract used in the analysis. The dataset enabled a comprehensive analysis of all hospital admissions for IID in England over the seven year period between 2009 and 2015. Since

hospitals submit data to NHS Digital to calculate payments owed for care that has been provided on a patient-by-patient basis, it is likely that HES data captures all admissions that occur in England (NHS Digital, 2016b). Additionally, data analysis was conducted at the LSOA level, which is a relatively small area (LSOAs contain an average of approximately 1600 people). This meant that for each age subgroup there were over 30,000 observations available to analyse, which increased the precision of the estimates. Since the study analysed hospital admissions in the population of England, it can be assumed that the results are generalisable.

The main limitation of this analysis related to the availability of data. Information on potential confounding and mediating factors was limited to data that were available from open source national datasets. Lack of detail within the ethnicity and long-term health problem variables precluded an in-depth analysis of the effects of these variables on the relationship between deprivation and the hospitalisation outcomes. Additionally, information on repeat admissions by the same individual and hospital identifiers were not available within the dataset, which precluded the investigation of clustering at the individual and hospital level. It is not known to what extent these factors may have influenced the results of the study. However previous studies investigating inequalities in the risks and consequences of IID using individual-level data, have found that inequalities persist even after accounting for recurrent IID (Adams et al., 2017; Rose et al., 2017).

There are also general drawbacks of using routinely collected data for research purposes. There can be a lack of standardisation in the data collection process because data is usually collected by numerous individuals, who may record data items differently (Kane et al., 2000). Furthermore, since data are not collected by trained researchers or for research purposes this might negatively affect data quality. However, the objectivity of routine data can also be of great benefit, since data are unlikely to be affected by biases relating to the impartiality of researchers collecting the data, or reporting biases (Kane et al., 2000).

Methodological limitations of ecological studies can include ecological bias, whereby associations present at the group-level are not apparent at the individual-level, possibly due to unmeasured confounding or measurement error (Greenland and Robins, 1994). However, because ecological studies are able to capture risk factors and exposures that operate at the community-level (Pearce, 2000), it could be argued that ecological studies are in fact more appropriate for the study of infectious diseases compared to individual-level studies. Additionally, measuring SES at the area-level may capture the socioeconomic characteristics of a neighbourhood as well as the characteristics of individual residents. Some studies have

shown that living in an economically deprived neighbourhood confers a small additional increased risk of premature mortality and long term illness, regardless of individual SES (Ross and Mirowsky, 2008; Van Lenthe, 2006; Pickett and Pearl, 2001). For Study 3, individual-level data were not available to analyse which precluded an investigation of the relationship between contextual and compositional socioeconomic factors on the risk of hospitalisation for IID, however this may be an interesting avenue for future research.

7.4 IMPLICATIONS FOR POLICY AND PRACTICE

Having discussed the findings of this thesis and critiqued the methods that have been used to generate the findings, I now move on to evaluate the implications of the findings for current policy and practice in the UK.

Policies and interventions relevant to GI infections in the UK have mainly focused on monitoring and reducing the risk of acquiring an infection. Examples of these policies include national surveillance programmes that monitor the incidence of GI infections across the UK, and a number of regulations relating to the supply of food that ensure food produced in the UK and imported is safe to eat (FSA, 2011; GOV.UK, 2010). Some interventions designed to reduce foodborne disease have been implemented nationally, such as the Food Hygiene Rating Scheme to inform consumers about hygiene standards in food establishments, and the Food Hygiene Campaign designed to improve public awareness about good food hygiene practice in the home (FSA, 2011). Additionally, a rotavirus vaccine was introduced to the national childhood immunisation programme in July 2013 (GOV.UK, 2013).

These policies and interventions may have reduced the risk of infection and subsequent burden of disease, for example laboratory reports of rotavirus infections declined by around 70% following the introduction of the rotavirus vaccine (PHE, 2017d). However, less is known about whether these policies have reduced inequalities in the risks or consequences of GI infections. Tackling inequalities in GI infections does not appear to be a priority area within current policy, and this may be because the evidence base has painted a confusing picture as to the extent of inequalities in the risk of infection in particular. The systematic review presented in this thesis attempted to address this problem, and the results suggest that in high income countries socioeconomically disadvantaged children are at greater risk of

symptomatic GI infections compared to their more affluent counterparts. Furthermore, the findings of this thesis suggest that inequalities exist in several consequences of GI infections, such as symptom severity, sickness absence, emergency hospital admissions and the duration of hospital admissions. These inequalities appear to affect both children and adults, and may exacerbate income inequalities due to the disproportionate impact of lost earnings on those in lower social positions. Policies and interventions designed to reduce the risk of acquiring an infection, are unlikely to adequately address the unequal consequences of GI infections that have been observed.

As discussed in this chapter, inequalities in consequences of GI infections such as the severity of illness, likely arise from socially patterned environmental, psychosocial and material factors that compromise immune functioning. This suggests that a broad approach is needed to tackle inequalities in the risks and consequences of GI infections; one that addresses the key social determinants of health and the factors that determine their unequal distribution across the social hierarchy (Graham, 2004). Diderichsen's model of the mechanisms of health inequality (Figure 7.2) highlights a number of entry points for policies that are designed to tackle inequalities in health (Diderichsen, Evans and Whitehead, 2001). These policy entry points occur throughout the pathway from reducing social stratification, to reducing differential exposures to health damaging risks, and reducing the differential consequences of being ill.

To reduce differential exposures to health damaging risks and subsequently reduce inequalities in health outcomes, the Strategic Review of Health Inequalities in England post-2010 (otherwise known as the Marmot Review) suggests action is required across all the social determinants of health (Marmot et al., 2010). The review recommends that policies should specifically address inequalities in early child development, in young people's educational achievement and acquisition of skills, in employment and working conditions, in housing and neighbourhood conditions, and in social and health services (Marmot et al., 2010). Additionally, action needs to be taken across the whole of society to not only alleviate the health damaging effects of poverty, but to also reduce the social gradients observed in health outcomes (Marmot, 2013). To achieve this, the review proposes that actions should be universal, but greater intensity of action is likely to be needed for those who are most disadvantaged (Marmot et al., 2010). These actions would likely help to reduce inequalities in GI infections, since as discussed, inequalities in GI infections likely arise from broader socially patterned factors that compromise immune functioning.

Communities, clinicians, public health bodies, health services and local and national government all have important roles to play in renewing efforts to deliver these policy objectives and reduce health inequalities. Starting with communities, the Marmot Review and a report prepared by the Inquiry Panel on Health Equity for the North of England (the Due North report) recommend that policies which enhance democratic engagement and empower individuals to take control over the way their communities are run, are central to reducing health inequalities (Whitehead et al., 2014; Marmot et al., 2010). Local government, councils and providers of public services could enable greater participation, so that disadvantaged communities have a larger influence in local decision-making and how public resources and community assets are used (Whitehead et al., 2014).

The British Medical Association (BMA) makes several suggestions of how doctors can advocate for the health needs of their patients and wider society, by lobbying local and national policy-makers and commissioners. Doctors can have a strong influence on policy decisions by highlighting how socioeconomic factors influence health and wellbeing, and arguing for action to tackle the social determinants of health (BMA, 2017; BMA, 2016). Additionally, doctors are in a unique position to directly support patients within the healthcare setting, for example by signposting patients to non-medical sources of support, such as welfare advice, financial advice services, food banks or community projects (BMA, 2017; BMA, 2016).

Similarly, the NHS and PHE have an essential role in advocating for action to reduce health inequalities. The Due North report makes a number of suggestions of how the health sector can do more to promote health equity (Whitehead et al., 2014). For example, the NHS has great potential to influence the social determinants of health as an employer. The NHS could lead the way in promoting high quality employment by improving working conditions and expanding training and apprenticeship programmes (Whitehead et al., 2014). Additionally, PHE could do more to independently lead and co-ordinate action on reducing health inequalities, and encourage all government departments to address health inequalities in all policies (Whitehead et al., 2014). Specifically in relation to GI infections, PHE could improve monitoring of inequalities in GI infections via their national routine surveillance systems. Continually measuring and monitoring the extent of inequalities in GI infections may enhance our understanding, and could assist in the evaluation of policies and interventions designed to tackle the problem.

Furthermore, the work in this thesis suggests that due consideration should be afforded to policies that address inequalities in the consequences of being ill with a GI infection.

Previous work has highlighted the importance of policies (such as income maintenance, labour market and vocational rehabilitation policies) in preventing the unequal consequences of being ill with a chronic condition or disability (Whitehead, Hanratty and Burström, 2009). Whilst the impact of being ill with a GI infection is likely to be less severe than being ill with a chronic condition, current employment law provides little protection for certain groups such as the self-employed when they require relatively short spells of sickness absence. Reforms to employment law regarding the self-employed may be of benefit. This would protect the rising number of individuals in self-employment in the UK (Field et al., 2017), granting them the right to receive the National Minimum Wage and Statutory Sick Pay, and would likely reduce the unequal consequences of falling ill with a GI infection.

Additionally, efforts to promote healthy development in early childhood, and to narrow the social gradient in educational attainment, as recommended within the Due North report and Marmot Review (Whitehead et al., 2014; Marmot et al., 2010), would likely help reduce the unequal burden of GI infections amongst children. For example, high quality interventions which support mothers to breastfeed might be of particular benefit in reducing both the incidence and severity of GI infections in children. Narrowing the educational attainment gap may also help to reduce the impact of sickness absence from school on educational outcomes.

Finally, as discussed in Chapter 2 of this thesis, socioeconomic stratification is a fundamental cause of health inequalities because a person's social position determines their exposure to health damaging risks, their vulnerability to ill health, and the consequences they experience as a result of illness. Policies that are focused 'upstream', that are designed to tackle the root causes of health inequalities (i.e. the level of socioeconomic stratification in a society), as well as the social determinants of health, are widely regarded by public health researchers as having the greatest potential for reducing health inequalities (Douglas, 2016; Smith, Bambra and Hill, 2016; Katikireddi et al., 2013; Graham, 2007). Establishing a minimum income for healthy living, and introducing more progressive taxation systems to redistribute wealth fairly and reduce levels of extreme income inequality, would likely improve social cohesion and the health of all members of society (Marmot, 2013; Marmot et al., 2010).

As Nelson Mandela once said:

> *"as long as poverty, injustice and gross inequality persist in our world, none of us can truly rest. (...) [P]overty is not natural. It is man-made and it can be overcome and eradicated by the actions of human beings."*
>
> (Mandela, 2005)

7.5 FUTURE RESEARCH

Several areas for further research have been highlighted in this discussion, especially in terms of investigating explanatory factors for the social gradients that have been observed. Gaining a better understanding of the mechanisms that explain the social gradient in emergency hospital admissions for IID is a research avenue that particularly warrants further investigation. For example, investigating whether factors related to primary care quality help to explain inequalities in admission rates, and assessing the extent to which various immune compromising factors explain inequalities in IID severity and how this relates to the social gradient in admission rates. Evidence from such studies may assist in the development of targeted interventions and policies to help reduce inequalities in this severe outcome. Additionally, the methods used in this thesis could be applied to other diseases that display social gradients in consequences such as hospitalisations, to examine whether inequalities in the risk of acquiring a disease help to explain inequalities in the consequences that follow.

The work presented in this thesis has been conducted as part of a wider research theme exploring inequalities in GI infections. Ongoing research is investigating how households with young children and their social and care networks, including friends, carers and childcare providers, manage a case of GI infection and its onward transmission and how this is shaped by the socioeconomic environment in which they live (Rotheram, 2017). This includes how households make decisions around engaging with healthcare services for GI infections. Additional qualitative research aims to understand how healthcare professionals make treatment, referral, follow-up and monitoring decisions related to GI infections, and whether there are differences in the lived experiences of households that have received formal healthcare for a GI infection in socioeconomically contrasting areas (McGarrol, 2017). It is hoped that this qualitative work, combined with the work within this thesis and an additional thesis exploring inequalities in the risk of infection, will appreciably improve understanding of inequalities in GI infections.

On a personal note, I feel that I have learned a great deal from writing this thesis and conducting the studies contained within it. My knowledge of both GI infections and inequalities in health has been enhanced, and I have gained skills in performing systematic reviews, analysing data and writing for publication. I will continue to use these skills and my understanding of health inequalities and GI infections in my future work. For example, I am currently using HES data to explore inequalities in hospital admissions for chronic conditions such as cardiovascular diseases, and to evaluate the impact of a local healthcare intervention on hospital admissions for respiratory conditions. Furthermore, planning is underway to investigate differences in the composition of the gut microbiota, amongst individuals with varying socioeconomic backgrounds.

Additionally, I intend to continue disseminating the work of this thesis via conferences and journal publications. Communicating the findings of my work to a wide range of audiences is an important aspect of this project. This is because the evidence within this thesis suggests that inequalities in the consequences of GI infections are present, and that advocating for action to reduce these inequalities is necessary. The publications that have already arisen from this thesis are presented in Appendix 7, and two additional journal articles are expected to be published in 2017–18.

7.6 Conclusions

This thesis has assessed the extent of socioeconomic inequalities in various consequences of GI infections, and has explored potential explanations for the inequalities identified. Firstly, a systematic literature review identified age as a statistically significant modifier of the association between SES and the risk of symptomatic GI infections in high income countries. Children (aged <18 years) of lower SES, but not adults, had a greater risk of infection compared to their more affluent counterparts. Secondly, analysis of the UK-based IID2 study revealed that IID cases aged ≥5 years, of lower SES, were more likely to experience severe symptoms and be absent from work or school. The association between SES and sickness absence was largely explained statistically by greater symptom severity amongst the more disadvantaged cases. Finally, an ecological analysis of English HES data showed that increasing neighbourhood deprivation was associated with increasing emergency hospital admission rates and admission duration for IID, for both adults and children. The socio-spatial gradient in admission rates was partly explained statistically by geographical factors

and the higher prevalence of long-term health problems in the more deprived neighbourhoods.

Important consequences of GI infections, such as sickness absence and emergency hospitalisation, incur heavy burdens for individuals and societies, and evidence from this thesis suggests these outcomes disproportionately affect disadvantaged groups. Inequalities in sickness absence can largely be explained by greater symptom severity amongst those of lower SES, and it seems highly plausible that differential disease severity by SES could help to explain the socio-spatial gradients in emergency hospitalisations for GI infections which have been observed across all age groups. Additionally, it seems unlikely that differential risk of infection by SES can explain the socio-spatial gradient in emergency hospital admissions for adults, but may contribute to inequalities in hospital admission rates for children.

Inequalities in the consequences of GI infections likely arise from socially patterned environmental, psychosocial and material factors that compromise immune functioning, such as low breastfeeding rates, comorbidity, smoking, increased levels of chronic stress and nutritional deficiencies. This suggests that a broad approach is needed to tackle inequalities in GI infections. Additionally, inequalities in consequences such as sickness absence and hospitalisation have the potential to exacerbate existing income inequalities due to the disproportionate impact of lost earnings on those in lower social positions. Policies and interventions that are designed to reduce the risk of acquiring an infection in general, are unlikely to adequately address the unequal consequences that have been observed. The work in this thesis suggests that more consideration should be afforded to policies and that address inequalities in GI infections, and tackling these inequalities will likely require action across the social determinants of health.

Appendices to Chapter 4

The appendix for Chapter 4 features supplementary material to support the information provided within the main chapters relevant to the systematic literature review.

A.4.1 DATABASE SEARCH TERMS

MEDLINE (Ovid)

#1. Exp Socioeconomic Factors/
#2. Education*.mp.
#3. Exp Employment/
#4. Income*.mp.
#5. Occupation*.mp.
#6. Poverty.mp.
#7. Poorest.mp.
#8. exp Social Class/
#9. Inequalit*.mp.
#10. Socioeconomic*.mp.
#11. Depriv*.mp.
#12. Disadvantag*.mp.
#13. Salary.mp.
#14. Underprivileged.mp.
#15. Social determinant*.mp.
#16. (Social adj1 factor*).mp
#17. Socio*.mp

#18. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17

#19. exp Norovirus/
#20. Acute gastroenteritis.mp.
#21. infectious intestinal disease*.mp.
#22. gastrointestinal infection*.mp.
#23. exp Diarrhea/
#24. Rotavirus.mp.
#25. gastrointestinal pathogen*.mp.
#26. gastrointestinal bacteria.mp.
#27. enteric infection*.mp.
#28. diarrh*.mp.
#29. stomach flu.mp.
#30. gastric flu.mp.
#31. stomach bug*.mp.
#32. stomach virus*.mp.
#33. Exp Campylobacter/
#34. Exp Escherichia coli/
#35. Enterobacteriaceae Infection*.mp.
#36. Dysentery, Bacillary.mp
#37. Exp Escherichia coli Infections/

#38. Yersinia enterocolitica.mp.
#39. Exp Salmonella Infections/
#40. Exp Cryptosporidiidae/
#41. Exp Salmonella/
#42. Exp Shigella/
#43. Exp Giardia/
#44. Escherichia coli.mp.
#45. Exp Listeria/
#46. Small round structured virus*.mp.
#47. Winter vomiting disease*.mp.
#48. Sapovirus.mp.
#49. Caliciviridae.mp.
#50. VTEC.mp.
#51. STEC.mp.
#52. exp Foodborne Diseases/
#53. Food poisoning*.mp.
#54. Scombro*.mp.
#55. Clostridium perfringens.mp.
#56. Bacillus cereus.mp.
#57. Hepatitis A.mp.
#58. Hepatitis E.mp.

#59. 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58

#60. exp Australia/
#61. exp Austria/
#62. exp Belgium/
#63. exp Canada/
#64. exp Chile/
#65. exp Czech Republic/
#66. exp Denmark/
#67. exp Estonia/
#68. exp Finland/
#69. exp France/
#70. exp Germany/
#71. exp Greece/
#72. exp Hungary/
#73. exp Iceland/
#74. exp Ireland/
#75. exp Israel/
#76. exp Italy/
#77. exp Japan/
#78. exp Korea/
#79. exp Luxembourg/
#80. exp Mexico/
#81. exp Netherlands/
#82. exp New Zealand/
#83. exp Norway/
#84. exp Poland/
#85. exp Portugal/
#86. exp Slovak Republic/
#87. exp Slovenia/
#88. exp Spain/
#89. exp Sweden/

#90. exp Switzerland/
#91. exp Turkey/
#92. exp United Kingdom/
#93. exp United States/

#94. 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93

#95. 18 and 59 and 94

Scopus - TITLE-ABS-KEY	Web of Science Core Collection - Topic
#1. "Career mobility"	
#2. Poverty	
#3. "Social class*"	
#4. "Social mobility"	
#5. Education*	
#6. Employment	
#7. Unemployment	
#8. Income*	
#9. Occupation*	
#10. Poor*	
#11. Inequalit*	
#12. Depriv*	
#13. Disadvantag*	
#14. Salary	
#15. Underprivileged	
#16. "Social determinant*"	
#17. Social pre/1 factor*	Social near/1 factor*
#18. Socio*	

#19. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18

#20. Norovirus
#21. "Norwalk virus"
#22. "Acute gastroenteritis"
#23. "Infectious intestinal disease*"
#24. "Gastrointestinal infection*"
#25. Rotavirus
#26. "Gastrointestinal pathogen*"
#27. "Gastrointestinal bacteria"
#28. "Enteric infection*"
#29. Diarrh*
#30. "Stomach flu"
#31. "Gastric flu"
#32. "Stomach bug*"
#33. "Stomach virus*"
#34. "Escherichia coli"
#35. "Enterobacteriaceae Infection*"
#36. Dysentery Bacillary
#37. "Yersinia enterocolitica"
#38. "paratyphoid fever"
#39. "typhoid fever"
#40. "Small round structured virus*"

#41. “Winter vomiting disease*”
#42. Sapovirus
#43. Caliciviridae
#44. Campylobacter*
#45. Cryptospor*
#46. Salmonell*
#47. Shigell*
#48. Giardia*
#49. Listeri*
#50. VTEC
#51. STEC
#52. “Foodborne Disease*”
#53. Botulism
#54. “Staphylococcal Food Poisoning*”
#55. “Food poisoning*”
#56. Scombro*
#57. “Clostridium perfringens”
#58. “Bacillus cereus”
#59. “Hepatitis A”
#60. “Hepatitis E”

#61. 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60

#62. Australia*
#63. “New South Wales”
#64. “Northern Territory”
#65. Queensland
#66. Tasmania
#67. Victoria
#68. Austria
#69. Belgium
#70. Canada
#71. Alberta
#72. “British Columbia”
#73. Manitoba
#74. “New Brunswick”
#75. “Newfoundland and Labrador”
#76. “Northwest Territories”
#77. “Nova Scotia”
#78. Nunavut
#79. Ontario
#80. “Prince Edward Island”
#81. Quebec
#82. Saskatchewan
#83. “Yukon Territory”
#84. Chile
#85. “Czech Republic”
#86. Denmark
#87. Greenland
#88. Estonia
#89. Finland
#90. France
#91. Paris
#92. Germany

#93. Berlin
#94. Greece
#95. Hungary
#96. Iceland
#97. Ireland
#98. Israel
#99. Italy
#100. Rome
#101. Sicily
#102. Japan
#103. Tokyo
#104. Korea
#105. Seoul
#106. Luxembourg
#107. Mexico
#108. Netherlands
#109. “New Zealand”
#110. Norway
#111. Svalbard
#112. Poland
#113. Portugal
#114. “Slovak Republic”
#115. Slovakia
#116. Slovenia
#117. Spain
#118. Sweden
#119. Switzerland
#120. Turkey
#121. “United Kingdom”
#122. “Great Britain”
#123. “Channel Islands”
#124. Guernsey
#125. England
#126. London
#127. Scotland
#128. Hebrides
#129. Wales
#130. “United States”
#131. “Appalachian Region”
#132. Alabama
#133. Georgia
#134. Kentucky
#135. Maryland
#136. “New York”
#137. Carolina
#138. Ohio
#139. Pennsylvania
#140. Tennessee
#141. Virginia
#142. “Great Lakes Region”
#143. Illinois
#144. Chicago
#145. Indiana
#146. Michigan
#147. Minnesota
#148. Wisconsin

#149. “Mid-Atlantic Region”
#150. Delaware
#151. “District of Columbia”
#152. Baltimore
#153. “New Jersey”
#154. Philadelphia
#155. Iowa
#156. Kansas
#157. Missouri
#158. Nebraska
#159. Dakota
#160. Oklahoma
#161. “New England”
#162. Connecticut
#163. Maine
#164. Massachusetts
#165. Boston
#166. “New Hampshire”
#167. “Rhode Island”
#168. Vermont
#169. Idaho
#170. Montana
#171. Oregon
#172. Washington
#173. Wyoming
#174. “Pacific States”
#175. Alaska
#176. California
#177. “Los Angeles”
#178. “San Francisco”
#179. Hawaii
#180. Arkansas
#181. Florida
#182. Louisiana
#183. “New Orleans”
#184. Mississippi
#185. Arizona
#186. Colorado
#187. Nevada
#188. “New Mexico”
#189. Texas
#190. Utah

#191. 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100 or 101 or 102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120 or 121 or 122 or 123 or 124 or 125 or 126 or 127 or 128 or 129 or 130 or 131 or 132 or 133 or 134 or 135 or 136 or 137 or 138 or 139 or 140 or 141 or 142 or 143 or 144 or 145 or 146 or 147 or 148 or 149 or 150 or 151 or 152 or 153 or 154 or 155 or 156 or 157 or 158 or 159 or 160 or 161 or 162 or 163 or 164 or 165 or 166 or 167 or 168 or 169 or 170 or 171 or 172 or 173 or 174 or 175 or 176 or 177 or 178 or 179 or 180 or 181 or 182 or 183 or 184 or 185 or 186 or 187 or 188 or 189 or 190

#192. 19 and 61 and 191

A.4.2 THE LQAT FORMS USED FOR QUALITY APPRAISAL

STUDY ID		EXTRACTED BY		EXTRACTION DATE	DD	MM	YY

GAPP: METHODOLOGICAL QUALITY APPRAISAL CROSS SECTIONAL STUDIES

PART A: Study Sample

QUALITY 1 – SELECTION BIAS

Is there evidence of selection bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

QUALITY 2 – RESPONSE BIAS

Is there evidence of response bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

PART B: Exposure Assessment SES

QUALITY 3 – Direct exposure

Ranking of exposure measurement?	
SES measured and analysed per group	
SES measured per individual based on aggregate SES data	*
SES measured and analysed per individual directly	**

QUALITY 4 – RECALL BIAS

Is there evidence of recall bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

QUALITY 5 – MEASUREMENT BIAS

Is there evidence of measurement bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

1.

PART C: Outcome Assessment

QUALITY 6 – Assessment of GI infection

Ranking of outcome assessment?	
Self-reported recall	
Self-reported prospectively	*
Health record/physician diagnosed via symptoms	**
Laboratory confirmation	***

QUALITY 7 – BIAS IN ASCERTAINMENT

Is there evidence of ascertainment bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

PART D: Analysis of Results

QUALITY 8 – ADJUSTMENT FOR CONFOUNDING

Is there adjustment for confounding?	
No	Limited or no adjustment for confounding.
Yes - adequate	The main confounders adjusted for: age, sex (*)
Yes - good	Majority of known confounders in model (**)

PART E: Qualifying comments

QUALITY 9: indicate overall assessment and specific issues that you would like to draw attention to:

Total Stars	/12

2.

STUDY ID		EXTRACTED BY		EXTRACTION DATE	DD	MM	YY

GAPP: METHODOLOGICAL QUALITY APPRAISAL CASE-CONTROL STUDIES

PART A: Study Sample

QUALITY 1 – CASE SELECTION BIAS

Is there evidence of selection bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

QUALITY 2 – CONTROL SELECTION BIAS

Is there evidence of response bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

PART B: Exposure Assessment SES

QUALITY 3 – Direct exposure

Ranking of exposure measurement?	
SES measured and analysed per group	
SES measured per individual based on aggregate SES data	*
SES measured and analysed per individual directly	**

QUALITY 4 – RECALL BIAS

Is there evidence of recall bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

QUALITY 5 – MEASUREMENT BIAS

Is there evidence of measurement bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

1.

PART C: Outcome Assessment

QUALITY 6 – Assessment of GI infection

Ranking of outcome assessment?	
Self-reported symptoms recall	
Self-reported symptoms prospectively	*
Health record/physician diagnosed via symptoms	**
Laboratory confirmation	***

PART D: Analysis of Results

QUALITY 7 – ADJUSTMENT FOR CONFOUNDING
(Including matching at design stage)

Is there adjustment for confounding?	
No	Limited or no adjustment for confounding.
Yes - adequate	The main confounders adjusted for: age, sex (*)
Yes - good	Majority of known confounders in model (**)

PART E: Qualifying comments

QUALITY 8: indicate overall assessment and specific issues that you would like to draw attention to:

Total Stars	/11

2.

STUDY ID		EXTRACTED BY		EXTRACTION DATE	DD	MM	YY

GAPP: METHODOLOGICAL QUALITY APPRAISAL COHORT STUDIES

PART A: Study Sample

QUALITY 1 – SELECTION BIAS

Is there evidence of selection bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

QUALITY 2 – RESPONSE BIAS

Is there evidence of response bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

QUALITY 3 – BIAS IN FOLLOW-UP

Is there evidence of bias in follow up?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

PART B: Exposure Assessment SES

QUALITY 4 – Direct exposure

Ranking of exposure measurement?	
SES measured and analysed per group	
SES measured per individual based on aggregate SES data	*
SES measured and analysed per individual directly	**

1.

QUALITY 5 – MEASUREMENT BIAS

Is there evidence of measurement bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

PART C: Outcome Assessment

QUALITY 6 – Assessment of GI infection

Ranking of outcome assessment?	
Self-reported recall	
Self-reported prospectively	*
Health record/physician diagnosed via symptoms	**
Laboratory confirmation	***

QUALITY 7 – BIAS IN ASCERTAINMENT

Is there evidence of ascertainment bias?	
Yes	
Possible	
No	*
If Yes/possible, provide details:	

PART D: Analysis of Results

QUALITY 8 – ADJUSTMENT FOR CONFOUNDING

Is there adjustment for confounding?	
No	Limited or no adjustment for confounding.
Yes - adequate	The main confounders adjusted for: age, sex (*)
Yes - good	Majority of known confounders in model (**)

PART E: Qualifying comments

QUALITY 9: indicate overall assessment and specific issues that you would like to draw attention to:

Total Stars	/12

2.

A.4.3 BIBLIOGRAPHY OF INCLUDED STUDIES

1. Adlam SB, Perera S, Lake RJ, et al. (2011) Acute gastrointestinal illness in New Zealand: a community study. *Epidemiology and Infection*, 139(2):302-8.
2. Arena C, Amoros JP, Vaillant V, et al. (2014) Acute diarrhea in adults consulting a general practitioner in France during winter: incidence, clinical characteristics, management and risk factors. *BMC Infectious Diseases*, 14:574.
3. Arsenault J, Berke O, Michel P, et al. (2012) Environmental and demographic risk factors for campylobacteriosis: Do various geographical scales tell the same story? *BMC Infectious Diseases*, 12:318.
4. Baker D, Taylor H, Henderson J (1998) Inequality in infant morbidity: Causes and consequences in England in the 1990s. *Journal of Epidemiology and Community Health*, 52(7):451-8.
5. Banatvala N, Cramp A, Jones IR, et al. (1999) Salmonellosis in North Thames (East), UK: associated risk factors. *Epidemiology and Infection*, (122):201-7.
6. Barros H, Lunet N (2003) Association between child-care and acute diarrhea: A study in Portuguese children. *Rev Saude Publica*, 37(5):603-08.
7. Beale N, Peart C, Kay H, et al. (2010) 'ALSPAC' infant morbidity and council tax band: Doctor consultations are higher in lower bands. *European Journal of Public Health*, 20(4):403-8.
8. Beaudry M, Dufour R, Marcoux S (1995) Relation between infant feeding and infections during the first six months of life. *The Journal of Pediatrics*, 126(2):191-7.
9. Bemis K, Marcus R, Hadler JL (2014) Socioeconomic status and campylobacteriosis, Connecticut, USA, 1999-2009. *Emerging Infectious Diseases*, 20(7):1240-2.
10. Bessell PR, Matthews L, Smith-Palmer A, et al. (2010) Geographic determinants of reported human Campylobacter infections in Scotland. *BMC Public Health*, 10:423.
11. Biering-Sørensen S, Søndergaard G, Vitting Andersen K, et al. (2012) Time trends in socio-economic factors and risk of hospitalisation with infectious diseases in pre-school children 1985-2004: a Danish register-based study. *Paediatric and Perinatal Epidemiology*, 26(3):226-35.
12. Bless PJ, Schmutz C, Suter K, et al. (2014) A tradition and an epidemic: determinants of the campylobacteriosis winter peak in Switzerland. *European Journal of Epidemiology*, 29(7):527-37.
13. Borgnolo G, Barbone F, Scornavacca G, et al. (1996) A case-control study of Salmonella gastrointestinal infection in Italian children. *Acta Paediatrica*, 85(7):804-8.
14. Bozkurt AI, Özgür S, Özçirpici B (1999) The prevalance of diseases with diarrhoea among children under five years old in different socioeconomic levels. *Tr J of Medical Sciences*, 29:561-5.
15. Bozkurt AI, Özgür S, Özçirpici B (2003) Association between household conditions and diarrheal diseases among children in Turkey: a cohort study. *Pediatrics International*, 45(4):443-51.
16. Britton E, Hales S, Venugopal K, et al. (2010) The impact of climate variability and change on cryptosporidiosis and giardiasis rates in New Zealand. *Journal of Water and Health*, 8(3):561-71.
17. Chang M, Groseclose SL, Zaidi AA, et al. (2009) An ecological analysis of sociodemographic factors associated with the incidence of salmonellosis, shigellosis, and E. coli O157:H7 infections in US counties. *Epidemiology and Infection*, 137(6):810-20.
18. Cohen SA, Egorov AI, Jagai JS, et al. (2008) The SEEDs of two gastrointestinal diseases: socioeconomic, environmental, and demographic factors related to cryptosporidiosis and giardiasis in Massachusetts. *Environmental Research*, 108(2):185-91.
19. Danis K, Di Renzi M, O'Neill W, et al. (2009) Risk factors for sporadic Campylobacter infection: an all-Ireland case-control study. *Euro Surveillance*, 14(7).
20. De Wit MA, Koopmans MP, Kortbeek LM, et al. (2001) Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. *American Journal of Epidemiology*, 154(7):666-74.
21. De Wit MA, Koopmans MP, van Duynhoven YT (2003) Risk Factors for Norovirus, Sapporo-like Virus, and Group A Rotavirus Gastroenteritis. *Emerging Infectious Diseases*, 9(12):1563-70.
22. Dennehy PH, Cortese MM, Begue RE, et al. (2006) A case-control study to determine risk factors for hospitalization for rotavirus gastroenteritis in U.S. children. *Pediatric Infectious Disease Journal*, 25(12):1123-31.
23. Doorduyn Y, Van Pelt W, Havelaar AH (2012) The burden of infectious intestinal disease (IID)

in the community: a survey of self-reported IID in The Netherlands. *Epidemiology and Infection*, 140(7):1185-92.

24. Doré K, Buxton J, Henry B, et al. (2004) Risk factors for Salmonella typhimurium DT104 and non-DT104 infection: A Canadian multi-provincial case-control study. *Epidemiology and Infection*, 132(3):485-93.

25. Duggirala HJ, Hassig SE, Santana S, et al. (2005) Evaluation of a hepatitis A immunization program. *Pediatric Infectious Disease Journal*, 24(11):974-8.

26. Eaton-Evans J, Dugdale AE (1987) Effects of feeding and social factors on diarrhoea and vomiting in infants. *Archives of Disease in Childhood*, 62(5):445-8.

27. Ethelberg S, Olesen B, Neimann J, et al. (2006) Risk factors for diarrhea among children in an industrialized country. *Epidemiology*, 17(1):24-30.

28. Etiler N, Velipasaoglu S, Aktekin M (2004) Risk factors for overall and persistent diarrhoea in infancy in Antalya, Turkey: A cohort study. *Public Health*, 118(1):62-9.

29. Evans MR, Sarvotham T, Thomas DR, et al. (2006) Domestic and travel-related foodborne gastrointestinal illness in a population health survey. *Epidemiology and Infection*, 134(4):686-93.

30. Faustini A, Marinacci C, Fabrizi E, et al. (2006) The impact of the Catholic Jubilee in 2000 on infectious diseases. A case-control study of giardiasis, Rome, Italy 2000-2001. *Epidemiology and Infection*, 134(3):649-58.

31. Fein SB, Lin CTJ, Levy AS (1995) Foodborne illness: Perceptions, experience, and preventive behaviors in the United States. *Journal of Food Protection*, 58(12):1405-11.

32. Fewtrell L, Kay D, Wyer M, et al. (1997) An investigation into the possible links between shigellosis and hepatitis A and public water supply disconnections. *Public Health*, 111(3):179-81.

33. Friedman CR, Hoekstra RM, Samuel M, et al. (2004) Risk factors for sporadic Campylobacter infection in the United States: A case-control study in FoodNet sites. *Clinical Infectious Diseases*, 38(Suppl 3):S285-96.

34. Fullerton KE, Ingram LA, Jones TF, et al. (2007) Sporadic campylobacter infection in infants: a population-based surveillance case-control study. *The Pediatric Infectious Disease Journal*, 26(1):19-24.

35. Gillespie IA, O'Brien SJ, Penman C, et al. (2008) Demographic determinants for Campylobacter infection in England and Wales: Implications for future epidemiological studies. *Epidemiology and Infection*, 136(12):1717-25.

36. Gillespie I, Mook P, Little C, et al. (2010) Human listeriosis in England, 2001-2007: association with neighbourhood deprivation. *Euro Surveillance*, 15(27):7–16.

37. Green CG, Krause DO, Wylie JL (2006) Spatial analysis of campylobacter infection in the Canadian province of Manitoba. *International Journal of Health Geographics*, 5:2.

38. Gupta A, Polyak CS, Bishop RD, et al. (2004) Laboratory-confirmed shigellosis in the United States, 1989-2002: epidemiologic trends and patterns. *Clinical Infectious Diseases*, 38(10):1372-7.

39. Hall GV, Kirk MD, Ashbolt R, et al. (2006) Frequency of infectious gastrointestinal illness in Australia, 2002: Regional, seasonal and demographic variation. *Epidemiology and Infection*, 134(1):111-8.

40. Herikstad H, Yang S, Van Gilder TJ, et al. (2002) A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996-7. *Epidemiology and Infection*, 129(1):9-17.

41. Hu W, Mengersen K, Tong S (2009) Spatial analysis of notified cryptosporidiosis infections in Brisbane, Australia. *Annals of Epidemiology*, 19(12):900-7.

42. Hu W, Mengersen K, Tong S (2010) Risk factor analysis and spatiotemporal CART model of cryptosporidiosis in Queensland, Australia. *BMC Infectious Diseases*, 10:311.

43. Hughes GJ, Gorton R (2015) Inequalities in the incidence of infectious disease in the North East of England: a population-based study. *Epidemiology and Infection*, 143(1):189-201.

44. Iacono G, Merolla R, D'Amico D, et al. (2005) Gastrointestinal symptoms in infancy: a population-based prospective study. *Digestive and Liver Disease*, 37(6):432-8.

45. Jackson R, Smith D, Tabnak F, et al. (2015) Disparities of shigellosis rates among California children by race/ethnicity and census tract poverty level, 2000-2010. *Pediatric Infectious Disease Journal*, 34(8):843-7.

46. Jalava K, Ollgren J, Eklund M, et al. (2011) Agricultural, socioeconomic and environmental variables as risks for human verotoxigenic Escherichia coli (VTEC) infection in Finland. *BMC Infectious Diseases*, 11:275.

47. Jones TF, McMillian MB, Scallan E, et al. (2007) A population-based estimate of the substantial burden of diarrhoeal disease in the United States, FoodNet; 1996-2003. *Epidemiology and Infection*, 135(2):293-301.

48. Kass PH, Farver TB, Beaumont JJ, et al. (1992) Disease determinants of sporadic salmonellosis in four northern California counties. A case-control study of older children and adults. *Annals of Epidemiology*, 2(5):683-96.

49. Kotloff KL, Wasserman SS, Steciak JY, et al. (1988) Acute diarrhea in Baltimore children attending an outpatient clinic. *Pediatric Infectious Disease Journal*, 7(11):753-9.

50. Kum-Nji P, Mangrem CL, Wells PJ, et al. (2009) Is environmental tobacco smoke exposure a risk factor for acute gastroenteritis in young children? *Clinical Pediatrics*, 48(7):756-62.

51. Kyle RG, Kukanova M, Campbell M, et al. (2011) Childhood disadvantage and emergency admission rates for common presentations in London: an exploratory analysis. *Archives of Disease in Childhood*, 96(3):221-6.

52. Lake IR, Harrison FC, Chalmers RM, et al. (2007) Case-control study of environmental and social factors influencing cryptosporidiosis. *European Journal of Epidemiology*, 22(11):805-11.

53. Lake IR, Nichols G, Harrison FC, et al. (2009) Using infectious intestinal disease surveillance data to explore illness aetiology; a cryptosporidiosis case study. *Health and Place*, 15(1):333-9.

54. Lal A, Baker MG, French NP, et al. (2012) The epidemiology of human salmonellosis in New Zealand, 1997-2008. *Epidemiology and Infection*, 140(9):1685-94.

55. Lee LA, Shapiro CN, Hargrett-Bean N, et al. (1991) Hyperendemic shigellosis in the United States: A review of surveillance data for 1967-1988. *Journal of Infectious Diseases*, 164(5):894-900.

56. Ludvigsson JF, et al. (2006) Epidemiological study of constipation and other gastrointestinal symptoms in 8000 children. *Acta Paediatrica*, 95(5):573-80.

57. MacRitchie LA, Hunter CJ, Strachan NJ (2013) A population-based exposure assessment of risk factors associated with gastrointestinal pathogens: a Campylobacter study. *Epidemiology and Infection*, 141(5):976-86.

58. Majowicz SE, Horrocks J, Bocking K (2007) Demographic determinants of acute gastrointestinal illness in Canada: a population study. *BMC Public Health*, 7:162.

59. McAteer A, Elliott AM, Hannaford PC (2011) Ascertaining the size of the symptom iceberg in a UK-wide community-based survey. *British Journal of General Practice*, 61(582):e1-e11.

60. McPherson M, Lalor K, Combs B, et al. (2009) Serogroup-specific risk factors for Shiga toxin-producing Escherichia coli infection in Australia. *Clinical Infectious Diseases*, 49(2):249-56.

61. Moorin RE, Heyworth JS, Forbes GM, et al. (2010) Long-term health risks for children and young adults after infective gastroenteritis. *Emerging Infectious Diseases*, 16(9):1440-7.

62. Neal KR, Slack RC (1997) Diabetes mellitus, anti-secretory drugs and other risk factors for campylobacter gastro-enteritis in adults: A case-control study. *Epidemiology and Infection*, 119(3):307-11.

63. Nichols GL, Richardson JF, Sheppard SK, et al. (2012) Campylobacter epidemiology: A descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. *BMJ Open*, 2(4): e001179.

64. Odoi A, Martin SW, Michel P, et al. (2004) Determinants of the geographical distribution of endemic giardiasis in Ontario, Canada: A spatial modelling approach. *Epidemiology and Infection*, 132(5):967-76.

65. Olowokure B, Hawker J, Weinberg J, et al. (1999) Deprivation and hospital admission for infectious intestinal diseases. *Lancet*, 353(9155):807-8.

66. Özkan S, Tüzün H, Görer N, et al. (2007) Water usage habits and the incidence of diarrhea in rural Ankara, Turkey. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(11):1131-5.

67. Özmert EN, Kilic M, Yurdakök K (2008) Environmental tobacco smoke: Is it a risk factor for diarrhea in 6-18 months old infants? *Central European Journal of Public Health*, 16(2):85-6.

68. Pardhan-Ali A, Wilson J, Edge VL, et al. (2013) Community-level risk factors for notifiable gastrointestinal illness in the Northwest Territories, Canada, 1991-2008. *BMC Public Health*, 13:63.

69. Pearl DL, Louie M, Chui L, et al. (2009) A multi-level approach for investigating socio-economic and agricultural risk factors associated with rates of reported cases of Escherichia coli O157 in humans in Alberta, Canada. *Zoonoses and Public Health*, 56(8):455-64.

70. Penrose AS, Wells EV, Aiello AE (2007) Infectious causation of chronic disease: Examining the relationship between Giardia lamblia infection and irritable bowel syndrome. *World Journal of*

Gastroenterology, 13(34):4574-8.

71. Phillips G, Tam CC, Rodrigues LC, et al. (2011) Risk factors for symptomatic and asymptomatic norovirus infection in the community. *Epidemiology and Infection*, 139(11):1676-86.

72. Pockett RD, Adlard N, Carroll S, et al. (2011) Paediatric hospital admissions for rotavirus gastroenteritis and infectious gastroenteritis of all causes in England: an analysis of correlation with deprivation. *Current Medical Research and Opinion*, 27(4):777-84.

73. Pollard CM, Meng X, Williamson S, et al. (2014) Eating out is associated with self-reported food poisoning: a Western Australia population perspective, 1998 to 2009. *Public Health Nutrition*, 17(10):2270-7.

74. Pyra M, Conover C, Howland J, et al. (2012) Determinants of campylobacteriosis notifications in New Zealand. *Epidemiology and Infection*, 140(11):2087-8.

75. Quigley MA, Cumberland P, Cowden JM, et al. (2006) How protective is breast feeding against diarrhoeal disease in infants in 1990s England? A case-control study. *Archives of Disease in Childhood*, 91(3):245-50.

76. Rind E, Pearce J (2010) The spatial distribution of campylobacteriosis in New Zealand, 1997-2005. *Epidemiology and Infection*, 138(10):1359-71.

77. Rodrigues LC, Cowden JM, Wheeler JG, et al. (2001) The study of infectious intestinal disease in England: Risk factors for cases of infectious intestinal disease with Campylobacter jejuni infection. *Epidemiology and Infection*, 127(2):185-93.

78. Sakuma M, Urashima M, Okabe N (2006) Verocytotoxin-producing Escherichia coli, Japan, 1999-2004. *Emerging Infectious Diseases*, 12(2):323-5.

79. Sargeant JM, Majowicz SE, Snelgrove J (2008) The burden of acute gastrointestinal illness in Ontario, Canada, 2005-2006. *Epidemiology and Infection*, 136(4):451-60.

80. Satterthwaite P, Pritchard K, Floyd D, et al. (1999) A case-control study of Yersinia enterocolitica infections in Auckland. *Australian and New Zealand Journal of Public Health*, 23(5):482-5.

81. Scallan E, Fitzgerald M, Collins C, et al. (2004) Acute gastroenteritis in northern Ireland and the Republic of Ireland: a telephone survey. *Communicable Disease and Public Health*, 7(1):61-7.

82. Seo JY, Seo JH, Kim MH, et al. (2012) Pattern of hepatitis A incidence according to area characteristics using national health insurance data. *Journal of Preventive Medicine and Public Health*, 45(3):164-73.

83. Seo JY, Choi BY, Ki M, et al. (2013) Risk factors for acute hepatitis a infection in Korea in 2007 and 2009: A case-control study. *Journal of Korean Medical Science*, 28(6):908-14.

84. Sethi D, Cumberland P, Hudson MJ, et al. (2001) A study of infectious intestinal disease in England: risk factors associated with group A rotavirus in children. *Epidemiology and Infection*, 126(1):63-70.

85. Simonsen J, Frisch M, Ethelberg S (2008) Socioeconomic risk factors for bacterial gastrointestinal infections. *Epidemiology*, 19(2):282-90.

86. Spencer SE, Marshall J, Pirie R, et al. (2012) The spatial and temporal determinants of campylobacteriosis notifications in New Zealand, 2001-2007. *Epidemiology and Infection*, 140(9):1663-77.

87. Stafford RJ, Schluter P, Kirk M, et al. (2007) A multi-centre prospective case-control study of campylobacter infection in persons aged 5 years and older in Australia. *Epidemiology and Infection*, 135(6):978-88.

88. Stone DH, Mitchell S, Packham B, et al. (1994) Prevalence and first-line treatment of diarrhoeal symptoms in the community. *Public Health*, 108(1):61-8.

89. Tam CC, Viviani L, Rodrigues LC, et al. (2013) The second study of infectious intestinal disease (IID2): Increased rates of recurrent diarrhoea in individuals aged 65 years and above. *BMC Public Health*, 13:739.

90. Teschke K, Bellack N, Shen H, et al. (2010) Water and sewage systems, socio-demographics, and duration of residence associated with endemic intestinal infectious diseases: a cohort study. *BMC Public Health*, 10:767.

91. Turkish Ministry of Health (1995) Turkey 1993: results from the demographic and health survey. *Studies in Family Planning*, 26(3):180-4.

92. Unicomb LE, Dalton CB, Gilbert GL, et al. (2008) Age-specific risk factors for sporadic Campylobacter infection in regional Australia. *Foodborne Pathogens and Disease*, 5(1):79-85.

93. Van Cauteren D, De Valk H, Vaux S, et al. (2012) Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiology and Infection*, 140(4):697-705.

94.	Varga C, Pearl DL, McEwen SA, et al. (2013) Evaluating area-level spatial clustering of Salmonella Enteritidis infections and their socioeconomic determinants in the greater Toronto area, Ontario, Canada (2007 - 2009): a retrospective population-based ecological study. *BMC Public Health*, 13:1078.
95.	Weisent J, Rohrbach B, Dunn JR, et al. (2012) Socioeconomic determinants of geographic disparities in campylobacteriosis risk: a comparison of global and local modeling approaches. *International Journal of Health Geographics*, 11(1):45.
96.	Whitney BM, Mainero C, Humes E, et al. (2015) Socioeconomic status and foodborne pathogens in Connecticut, USA, 2000–2011. *Emerging Infectious Diseases*, 21(9):1617-24.
97.	Wilking H, Hohle M, Velasco E, et al. (2012) Ecological analysis of social risk factors for Rotavirus infections in Berlin, Germany, 2007-2009. *International Journal of Health Geographics*, 11:37.
98.	Wilking H, Spitznagel H, Werber D, et al. (2013) Acute gastrointestinal illness in adults in Germany: a population-based telephone survey. *Epidemiology and Infection*, 141(11):2365-75.
99.	Xu Z, Hu W, Tong S (2015) The geographical co-distribution and socio-ecological drivers of childhood pneumonia and diarrhoea in Queensland, Australia. *Epidemiology and Infection*, 143(5):1096-104.
100.	Younus M, Hartwick E, Siddiqi AA, et al. (2007) The role of neighborhood level socioeconomic characteristics in Salmonella infections in Michigan (1997-2007): assessment using geographic information system. *International Journal of Health Geographics*, 6:56.
101.	Younus M, Wilkins MJ, Davies HD, et al. (2010) The role of exposures to animals and other risk factors in sporadic, non-typhoidal Salmonella infections in Michigan children. *Zoonoses and Public Health*, 57(7-8):e170-6.
102.	Zappe Pasturel B, Cruz-Cano R, Rosenberg Goldstein RE, et al. (2013) Impact of rurality, broiler operations, and community socioeconomic factors on the risk of campylobacteriosis in Maryland. *American Journal of Public Health*, 103(12):2267-75.

A.4.4 CHARACTERISTICS OF INCLUDED STUDIES

Table A.4.1 Characteristics of included studies

Ref	First author	Year	Quality	Country/ Region	Age group	Level of analysis	Sample size	Study design	Pathogen/symptom	GI infection measure	SES measure	Reason for exclusion from meta-analysis
1	Adlam	2011	High	New Zealand	All	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Income	
2	Arena	2014	Medium	France	Adults	Individual	<200	Case-control	Multi-pathogen	GP Presentation	Multiple measures	
3	Arsenault	2012	Low	Canada	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Education	
4	Baker	1998	Low	UK/Ireland	Children	Individual	5001-10,000	Cohort	Acute GI infection (syndromic)	Population based survey	Education	
5	Banatvala	1999	Low	UK/Ireland	All	Area	>100,000	Ecological	Salmonellosis	Laboratory records	Deprivation	
6	Barros	2003	Low	Portugal	Children	Individual	200-1000	Cohort	Acute GI infection (syndromic)	Population based survey	Education	
7	Beale	2010	Low	UK/Ireland	Children	Area	5001-10,000	Cohort	Acute GI infection (syndromic)	Population based survey	Social class	Double counting cases
8	Beaudry	1995	Low	Canada	Children	Individual	200-1000	Cohort	Acute GI infection (syndromic)	Population based survey	Social class	
9	Bemis	2014	Low	United States	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Deprivation	
10	Bessell	2010	Low	UK/Ireland	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Deprivation	
11	Biering-Sorensen	2012	High	Denmark	Children	Individual	>100,000	Cohort	Acute GI infection (syndromic)	Hospital admission	Education	
12	Bless	2014	High	Switzerland	All	Individual	200-1000	Case-control	Campylobacteriosis	Laboratory records	Education	
13	Borgnolo	1996	High	Italy	Children	Individual	200-1000	Case-control	Salmonellosis	Hospital admission	Occupation	
14	Bozkurt	1999	Low	Turkey	Children	Individual	200-1000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Multiple measures	
15	Bozkurt	2003	Low	Turkey	Children	Individual	200-1000	Cohort	Acute GI infection (syndromic)	Population based survey	Multiple measures	
16	Britton	2010	Low	New Zealand	Not specified	Area	>100,000	Ecological	Multi-pathogen	Laboratory records	Deprivation	
17	Chang	2009	High	United States	All	Area	>100,000	Ecological	Multi-pathogen	Laboratory records	Multiple measures	Did not use a dichotomous outcome

18	Cohen	2008	Low	United States	All	Area	>100,000	Ecological	Multi-pathogen	Laboratory records	Income	
19	Danis	2009	Low	UK/Ireland	All	Individual	200-1000	Case-control	Campylobacteriosis	Laboratory records	Employment	
20	de Wit	2001	Medium	Netherlands	All	Individual	1001-5000	Cohort	Acute GI infection (syndromic)	Population based survey	Education	
21	de Wit	2003	Low	Netherlands	All	Individual	200-1000	Case-control	Multi-pathogen	Population based survey	Education	Double counting cases
22	Dennehy	2006	High	United States	Children	Individual	1001-5000	Case-control	Rotavirus	Hospital admission	Education	
23	Doorduyn	2012	High	Netherlands	All	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Education	
24	Doré	2004	High	Canada	All	Individual	200-1000	Case-control	Salmonellosis	Laboratory records	Education	
25	Duggirala	2005	Low	United States	All	Individual	200-1000	Case-control	Hepatitis A	Laboratory records	Education	
26	Eaton-Evans	1987	Low	Australia	Children	Individual	<200	Cohort	Acute GI infection (syndromic)	Population based survey	Occupation	Insufficient quantitative data
27	Ethelberg	2006	Medium	Denmark	Children	Individual	1001-5000	Case-control	Acute GI infection (syndromic)	Laboratory records	Multiple measures	
28	Etiler	2004	Low	Turkey	Children	Individual	200-1000	Cohort	Acute GI infection (syndromic)	Population based survey	Multiple measures	
29	Evans	2006	Medium	UK/Ireland	Adults	Individual	10,001-100,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Employment	
30	Faustini	2006	Medium	Italy	All	Individual	<200	Case-control	Giardiasis	Laboratory records	Education	
31	Fein	1995	High	United States	Adults	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Education	
32	Fewtrell	1997	Low	UK/Ireland	All	Area	>100,000	Ecological	Multi-pathogen	Laboratory records	Employment	Did not use a dichotomous outcome
33	Friedman	2004	Low	United States	All	Individual	1001-5000	Case-control	Campylobacteriosis	Laboratory records	Multiple measures	
34	Fullerton	2007	Medium	United States	Children	Individual	1001-5000	Case-control	Campylobacteriosis	Laboratory records	Multiple measures	
35	Gillespie	2008	Low	UK/Ireland	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Occupation	
36	Gillespie	2010	Low	UK/Ireland	All	Area	>100,000	Ecological	Listeriosis	Laboratory records	Deprivation	
37	Green	2006	High	Canada	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Social class	
38	Gupta	2004	Low	United States	All	Area	>100,000	Ecological	Shigellosis	Laboratory records	Deprivation	Insufficient quantitative data
39	Hall	2006	High	Australia	All	Individual	5001-10,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Multiple measures	
40	Herikstad	2002	Medium	United States	All	Individual	5001-10,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Multiple measures	Double counting cases

41	Hu	2009	Low	Australia	Not specified	Area	>100,000	Ecological	Cryptosporidiosis	Laboratory records	Multiple measures	
42	Hu	2010	Low	Australia	Not specified	Area	>100,000	Ecological	Cryptosporidiosis	Laboratory records	Social class	Did not use a dichotomous outcome
43	Hughes	2015	Low	UK/Ireland	All	Area	>100,000	Ecological	Multi-pathogen	Laboratory records	Deprivation	
44	Iacono	2005	High	Italy	Children	Individual	1001-5000	Cohort	Acute GI infection (syndromic)	Population based survey	Education	Insufficient quantitative data
45	Jackson	2015	Low	United States	Children	Area	>100,000	Ecological	Shigellosis	Laboratory records	Deprivation	
46	Jalava	2011	Low	Finland	All	Area	>100,000	Ecological	STEC	Laboratory records	Multiple measures	
47	Jones	2007	Medium	United States	All	Individual	10,001-100,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Education	
48	Kass	1992	Low	United States	All	Individual	200-1000	Case-control	Salmonellosis	Laboratory records	Income	
49	Kotloff	1988	Medium	United States	Children	Individual	200-1000	Case-control	Acute GI infection (syndromic)	Hospital admission	Education	Insufficient quantitative data
50	Kum-Nji	2009	High	United States	Children	Individual	200-1000	Cohort	Acute GI infection (syndromic)	Hospital admission	Employment	
51	Kyle	2011	Low	UK/Ireland	Children	Area	>100,000	Ecological	Acute GI infection (syndromic)	Hospital admission	Deprivation	Did not use a dichotomous outcome
52	Lake	2007	Low	UK/Ireland	All	Area	5001-10,000	Ecological	Cryptosporidiosis	Laboratory records	Occupation	
53	Lake	2009	Low	UK/Ireland	All	Area	>100,000	Ecological	Cryptosporidiosis	Laboratory records	Social class	Insufficient quantitative data
54	Lal	2012	Medium	New Zealand	All	Area	>100,000	Ecological	Salmonellosis	Multiple measures	Deprivation	
55	Lee	1991	Low	United States	All	Area	>100,000	Ecological	Shigellosis	Laboratory records	Deprivation	Insufficient quantitative data
56	Ludvigsson	2006	Low	Sweden	Children	Individual	5001-10,000	Cohort	Acute GI infection (syndromic)	Population based survey	Education	
57	MacRitchie	2013	Low	UK/Ireland	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Deprivation	Insufficient quantitative data
58	Majowicz	2007	Medium	Canada	All	Individual	5001-10,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Multiple measures	
59	McAteer	2011	High	UK/Ireland	Adults	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Multiple measures	Insufficient quantitative data
60	McPherson	2009	Medium	Australia	All	Individual	200-1000	Case-control	STEC	Laboratory records	Education	
61	Moorin	2010	Medium	Australia	All	Area	>100,000	Cohort	Acute GI infection (syndromic)	Hospital admission	Social class	
62	Neal	1997	Low	UK/Ireland	Adults	Individual	200-1000	Case-control	Campylobacteriosis	Laboratory records	Social class	

63	Nichols	2012	Low	UK/Ireland	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Deprivation	Insufficient quantitative data
64	Odoi	2004	Medium	Canada	Not specified	Area	>100,000	Ecological	Giardiasis	Laboratory records	Income	
65	Olowokure	1999	Low	UK/Ireland	All	Area	>100,000	Ecological	Acute GI infection (syndromic)	Hospital admission	Deprivation	
66	Özkan	2007	Low	Turkey	All	Individual	200-1000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Income	
67	Özmert	2008	Low	Turkey	Children	Individual	200-1000	Case-control	Acute GI infection (syndromic)	Hospital admission	Education	
68	Pardhan-Ali	2013	Low	Canada	All	Area	10,001-100,000	Ecological	Multi-pathogen	Laboratory records	Multiple measures	Insufficient quantitative data
69	Pearl	2009	Low	Canada	All	Area	>100,000	Ecological	STEC	Laboratory records	Income	Insufficient quantitative data
70	Penrose	2007	Low	United States	All	Area	>100,000	Ecological	Giardiasis	Laboratory records	Income	Insufficient quantitative data
71	Phillips	2011	Low	UK/Ireland	Children	Individual	200-1000	case-control	Norovirus	GP Presentation	Occupation	
72	Pockett	2011	Medium	UK/Ireland	Children	Area	>100,000	Ecological	Acute GI infection (syndromic)	Hospital admission	Deprivation	Did not use a dichotomous outcome
73	Pollard	2014	High	Australia	Adults	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Education	
74	Pyra	2012	Low	United States	Adults	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Multiple measures	
75	Quigley	2006	Medium	UK/Ireland	Children	Individual	200-1000	Case-control	Acute GI infection (syndromic)	GP Presentation	Occupation	Double counting cases
76	Rind	2010	High	New Zealand	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Social class	Did not use a dichotomous outcome
77	Rodrigues	2001	Medium	UK/Ireland	All	Individual	200-1000	Case-control	Campylobacteriosis	GP Presentation	Employment	
78	Sakuma	2006	Low	Japan	All	Area	>100,000	Ecological	STEC	Laboratory records	Income	Insufficient quantitative data
79	Sargeant	2008	Medium	Canada	All	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Multiple measures	
80	Satterthwaite	1999	Medium	New Zealand	All	Individual	200-1000	Case-control	Yersinia enterocolitica	Laboratory records	Education	
81	Scallan	2004	High	UK/Ireland	All	Individual	5001-10,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Occupation	
82	Seo	2012	High	Korea	Not specified	Area	>100,000	Ecological	Hepatitis A	Laboratory records	Multiple measures	Insufficient quantitative data
83	Seo	2013	Medium	Korea	All	Individual	1001-5000	Case-control	Hepatitis A	Hospital admission	Multiple measures	
84	Sethi	2001	Medium	UK/Ireland	Children	Individual	200-1000	Case-control	Rotavirus	GP Presentation	Occupation	

85	Simonsen	2008	Medium	Denmark	All	Individual	>100,000	Cohort	Multi-pathogen	Laboratory records	Multiple measures	
86	Spencer	2012	Low	New Zealand	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Deprivation	
87	Stafford	1996	Low	Australia	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Social class	
88	Stone	1994	Low	UK/Ireland	All	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Employment	Insufficient quantitative data
89	Tam	2013	High	UK/Ireland	All	Individual	5001-10,000	Cohort	Acute GI infection (syndromic)	Population based survey	Multiple measures	
90	Teschke	2010	Medium	Canada	All	Area	>100,000	Cohort	Acute GI infection (syndromic)	Multiple measures	Income	
91	Turkish Ministry of Health	1995	Low	Turkey	Children	Individual	1001-5000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Education	
92	Unicomb	2008	Medium	Australia	All	Individual	200-1000	Case-control	Campylobacteriosis	Laboratory records	Multiple measures	
93	Van Cauteren	2012	Medium	France	All	Individual	10,001-100,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Multiple measures	
94	Varga	2013	Medium	Canada	All	Area	>100,000	Ecological	Salmonellosis	Laboratory records	Multiple measures	
95	Weisent	2012	Low	United States	Not specified	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Multiple measures	
96	Whitney	2015	Low	United States	All	Area	>100,000	Ecological	Multi-pathogen	Laboratory records	Deprivation	
97	Wilking	2012	Low	Germany	All	Area	>100,000	Ecological	Rotavirus	Hospital admission	Employment	
98	Wilking	2013	Medium	Germany	Adults	Individual	10,001-100,000	Cross-sectional	Acute GI infection (syndromic)	Population based survey	Income	
99	Xu	2015	Low	Australia	Children	Area	>100,000	Ecological	Acute GI infection (syndromic)	Hospital admission	Social class	
100	Younus	2007	Low	United States	All	Area	>100,000	Ecological	Salmonellosis	Laboratory records	Multiple measures	
101	Younus	2010	Low	United States	Children	Individual	200-1000	Case-control	Salmonellosis	Laboratory records	Multiple measures	
102	Zappe Pasturel	2013	Low	United States	All	Area	>100,000	Ecological	Campylobacteriosis	Laboratory records	Multiple measures	

A.4.5 STUDIES WHICH ANALYSED THE SAME CASES

As detailed in the table above (A.4.1), four studies were excluded from the meta-analysis because they analysed the same cases as other studies. Including these studies would have meant that cases were double counted. These studies are described in Table A.4.2 below.

Table A.4.2 Studies which analysed the same cases

• De Wit et al. (2001a) • De Wit et al. (2003)	De Wit et al. (2001a) and De Wit et al. (2003) both analysed data from the Sensor study. De Wit et al. (2001a) analysed syndromic GI infections within the cohort study component, and De Wit et al. (2003) analysed norovirus and rotavirus infections within the nested case-control component. Since De Wit et al. (2003) analysed a smaller subset of data, only De Wit et al. (2001a) was included in the meta-analysis to avoid the 'double counting' of cases.
• Herikstad et al. (2002) • Jones et al. (2007)	Herikstad et al. (2002) and Jones et al. (2007) analysed different SES exposures, however they both analysed cases identified via a FoodNet population-based telephone survey conducted in the USA. Herikstad et al. (2002) analysed individuals surveyed between 1996–7, and Jones et al. (2007) analysed these same individuals, in addition to individuals surveyed in 1998–9 and 2000–3. Since Herikstad et al. (2002) analysed a smaller subset of data compared to Jones et al. (2007), the former was not included in the meta-analysis.
• Baker, Taylor and Henderson (1998) • Beale et al. (2010)	Baker, Taylor and Henderson (1998) and Beale et al. (2010) analysed different SES exposures (education level and council tax band, respectively), however both analysed data from the ALSPAC study. Only Baker, Taylor and Henderson (1998) was included in the meta-analysis since this study measured SES by education level which was the most commonly used SES measure amongst all studies included in the review. The results from both studies showed similar trends.
• Phillips et al. (2011) • Rodrigues et al. (2001) • Sethi et al. (2001) • Quigley et al. (2006)	Phillips et al. (2011), Rodrigues et al. (2001), Sethi et al. (2001) and Quigley et al. (2006) analysed data from the IID1 study in England, however they analysed different outcomes: norovirus, campylobacteriosis, rotavirus and syndromic GI infections, respectively. By analysing syndromic GI infections, Quigley et al. (2006) may have analysed the same cases as the other studies, and so this study was excluded from the meta-analysis.

Appendices to Chapter 5

The appendix for Chapter 5 features exploratory analyses, investigations of model assumptions and sensitivity analyses that were conducted to assess the robustness of the main results from the analysis of the IID2 study data presented in Chapter 5.

A.5.1 Exploratory analysis using GAMs

GAMs were used to visually assess the relationships between the continuous independent variable age, and the scaled log-odds of the IID symptom severity and sickness absence outcomes.

Figure A.5.1 shows the relationship between age and the symptom severity outcome. The Y axis in the plot shows the scaled log-odds of symptom severity. The scaled log-odds of symptom severity increase as age increases from five to around 20 years of age. Thereafter, the scaled log-odds of symptom severity tend to decrease with advancing age. Due to the non-linear relationship between age and symptom severity, a categorical age group variable (with categories: 5–14; 15–24; 25–44; 45–64 and 65+ years) was included when modeling the symptom severity outcome. The boundaries of the categories were chosen to demonstrate the relationship between age and symptom severity as displayed in Figure A.5.1.

Figure A.5.1 GAM showing the shape of the relationship between age and IID symptom severity

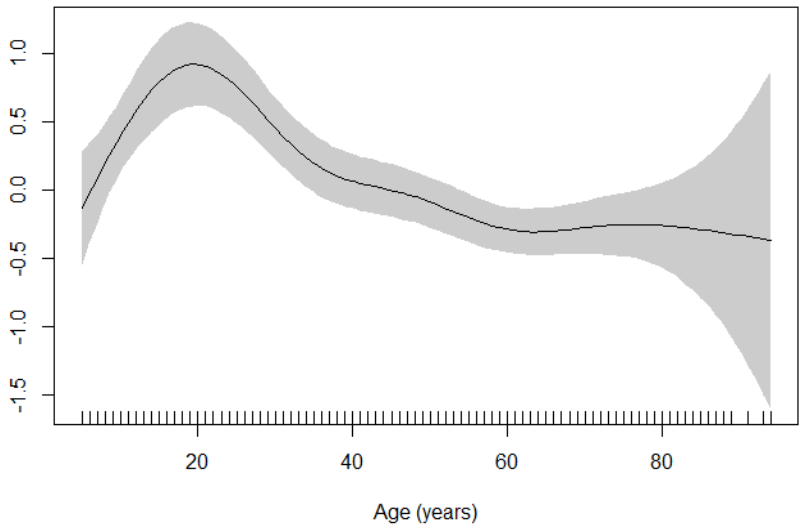
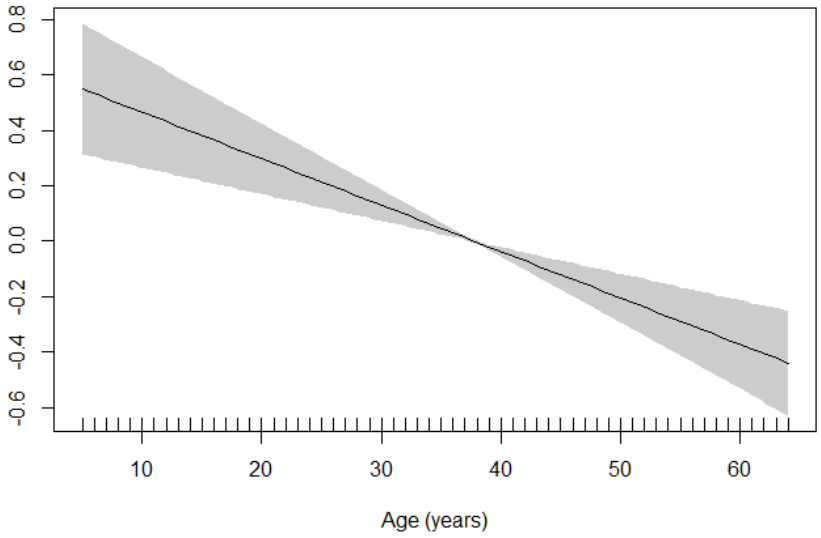


Figure A.5.2 shows the relationship between age and the sickness absence outcome. The Y axis in the plot shows the scaled log-odds of sickness absence. As can be seen, the scaled log-odds of sickness absence decrease with advancing age, in a linear fashion. The age variable was thus included as a continuous variable when modeling the sickness absence outcome.

Figure A.5.2 GAM showing the shape of the relationship between age and sickness absence due to IID



A.5.2 COMBINING CASES FROM COHORT AND GP STUDIES

To evaluate whether it was appropriate to combine individuals with IID from the Cohort and the GP Presentation study components of the IID2 study, an interaction term was added to the most parsimonious multivariate models from the main analysis for each outcome of interest (symptom severity and sickness absence). This interaction term, between NS-SEC (the primary exposure of interest) and study type (Cohort versus GP Presentation), was added to the models to test the null hypothesis that the relationships between NS-SEC and the outcomes did not differ between the Cohort and GP Presentation studies.

Table A.5.1 shows the results of multivariate ordinal logistic regression for IID symptom severity. The estimates for the NS-SEC terms can be interpreted as the effect of NS-SEC on symptom severity for the GP Presentation study. The estimates for the interaction term indicate the difference in the effect of NS-SEC on symptom severity between the GP

Presentation study and the Cohort study. The statistical significance of this difference is indicated by the p-values for the interaction term. The interaction term between NS-SEC and study type, was not statistically significant (at the <0.05 significance level) in the model, indicating that the relationship between NS-SEC and symptom severity was not significantly different between the Cohort and GP Presentation studies.

Table A.5.1 Multivariate model for severe IID symptoms versus mild or moderate symptoms combined for cases ≥5 years of age, with interaction term between NS-SEC and study type

	Odds ratio (95% CI)	p-value
Age group 15–24 years [a]	1.49 (0.79–2.87)	0.22166
Age group 25–44 years [a]	0.70 (0.46–1.06)	0.09449
Age group 45–64 years [a]	0.66 (0.44–0.98)	0.04078
Age group 65+ years [a]	0.42 (0.27–0.65)	<0.0001
Sex Male [b]	0.60 (0.47–0.76)	<0.0001
Ethnicity Non-White [c]	1.23 (0.66–2.33)	0.51992
NS-SEC Intermediate [d]	0.81 (0.54–1.21)	0.30585
NS-SEC Routine/manual [d]	1.39 (0.97–2.01)	0.07591
Study type Cohort [e]	0.08 (0.06–0.12)	<0.0001
Interaction term (NS-SEC Intermediate:Cohort)	**1.23 (0.66–2.25)†**	**0.51123**
Interaction term (NS-SEC Routine/manual:Cohort)	**0.94 (0.50–1.73)†**	**0.83392**

Number: 1164 IID cases included in model
CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification
[a] reference category = Age group 5–14 years
[b] reference category = Sex Female
[c] reference category = Ethnicity White
[d] reference category = NS-SEC Managerial/professional occupations
[e] reference category = Study type GP presentation
† Ratio of odds ratios

Table A.5.2 shows the results of multivariate logistic regression for sickness absence. The estimates for the NS-SEC terms can be interpreted as the effect of NS-SEC on absence for the GP Presentation study. The estimates for the interaction term indicate the difference in the effect of NS-SEC on sickness absence between the GP Presentation study and the Cohort study. The statistical significance of this difference is indicated by the p-values for the interaction term. The interaction term between NS-SEC and study type, was not statistically significant in the model, indicating that the relationship between NS-SEC and sickness absence was not significantly different between the Cohort and GP Presentation studies.

Table A.5.2 Multivariate model for sickness absence due to IID for cases of school/working age, with interaction term between NS-SEC and study type

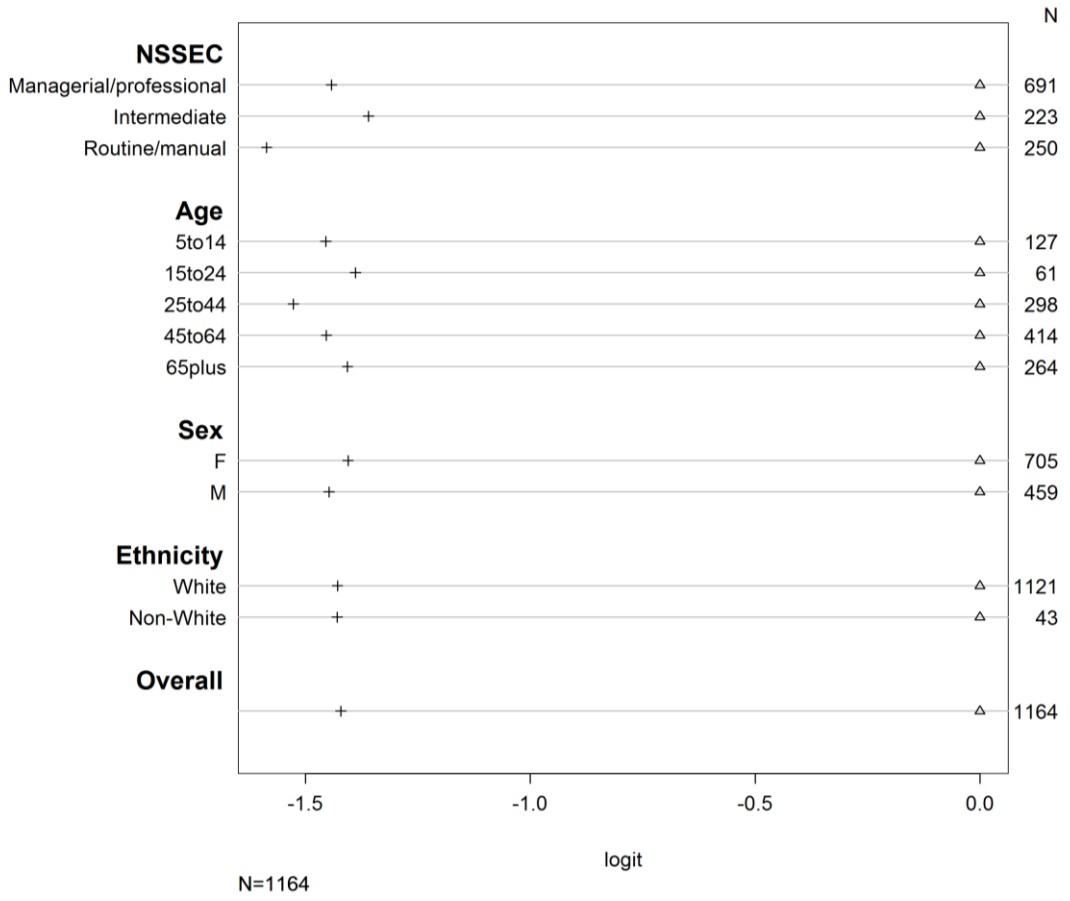
	Odds ratio (95% CI)	p-value
Age (years)	0.99 (0.98–1.00)	0.01267
Sex Male [a]	0.86 (0.63–1.19)	0.37066
Ethnicity Non-White [b]	1.76 (0.74–4.90)	0.23599
NS-SEC Intermediate [c]	0.91 (0.51–1.65)	0.75240
NS-SEC Routine/manual [c]	1.26 (0.75–2.16)	0.38271
Symptom severity Moderate [d]	3.13 (2.16–4.57)	<0.0001
Symptom severity Severe [d]	4.02 (2.49–6.54)	<0.0001
Study type Cohort [e]	0.65 (0.41–1.02)	0.05991
Interaction term (NS-SEC Intermediate:Cohort)	**1.24 (0.55–2.80)†**	**0.60242**
Interaction term (NS-SEC Routine/manual:Cohort)	**1.03 (0.44–2.40)†**	**0.94993**

Number: 818 IID cases included in model
CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification
[a] reference category = Sex Female
[b] reference category = Ethnicity White
[c] reference category = NS-SEC Managerial/professional occupations
[d] reference category = Symptom severity Mild
[e] reference category = Study type GP presentation
† Ratio of odds ratios

A.5.3 PROPORTIONAL ODDS ASSUMPTION

An assumption of ordinal logistic regression (used to mode the IID symptom severity outcome) is that the coefficients describing the relationship between each pair of outcome categories are the same. A graphical method was used to test this assumption for the most parsimonious multivariate model from the main analysis for the symptom severity outcome. To test the assumption, the dependent symptom severity variable was relabelled '1', '2' and '3' for the categories mild, moderate and severe, respectively. Predicted logits from univariate logistic regressions with the outcome symptom severity defined as either $\geq$2 or $\geq$3 are displayed graphically in Figure A.5.3. The proportional odds assumption holds if the differences between predicted logits for varying levels of the exposure variables are similar whether the outcome is defined by $\geq$2 or $\geq$3 (UCLA: Statistical Consulting Group, no date-a). In Figure A.5.3, the predicted logits from the regressions with symptom severity defined as $\geq$2 were normalised to zero. As can be seen, for each level of the exposure variable, the difference between predicted logits for each definition of the dependent variable were approximately similar, indicating the proportional odds assumption holds true, and therefore ordinal logistic regression was an appropriate method to use in this situation.

Figure A.5.3 Assessing the proportional odds assumption



A.5.4 CASES OF ALL AGES AND STRATIFIED RESULTS BY AGE

Multivariate analyses using the most parsimonious models for the symptom severity and sickness absence outcomes identified in the main analysis were repeated without applying age restrictions to the sample. Results stratified by children and adult age groups with the same lower and upper age limits as utilised in the main analysis are also presented. The results are displayed in Tables A.5.3, A.5.4 and A.5.5. As can be seen, the results confirmed those from the main analysis, although the number of child cases available to analyse was small. Children (aged ≥5 to <16 years) and adults whose main household earner was in a routine/manual compared to managerial/professional occupation, had greater odds of experiencing severe symptoms and sickness absence due to IID (Tables A.5.3 and A.5.4). These findings were also apparent when the outcomes were analysed using cases of all ages (0 to 90+ years) combined, however the association did not reach statistical significance when analysing the sickness absence outcome, which may have been related to measurement

error when assessing sickness absence in young children aged <5 years. Increasing symptom severity was associated with greater odds of sickness absence for children, adults and all ages combined (Table A.5.5).

Table A.5.3 Multivariate models for severe IID symptoms versus mild or moderate symptoms combined for cases of all ages and stratified by child and adult age groups

	All ages **0 to 90+ years** OR (95% CI)	**Children** **≥5 to <16 years** OR (95% CI)	**Adults** **≥16 years** OR (95% CI)
Age (years)	0.99 (0.99–1.00)	1.16 (1.02–1.31)	0.98 (0.97–0.99)
Sex			
Female	reference	reference	reference
Male	0.88 (0.72–1.09)	1.13 (0.58–2.20)	0.91 (0.72–1.15)
Ethnicity			
White	reference	reference	reference
Non-White	1.76 (1.04–2.99)	3.26 (1.04–10.77)	1.34 (0.68–2.68)
NS-SEC			
Managerial/ professional	reference	reference	reference
Intermediate	1.22 (0.93–1.59)	0.82 (0.31–2.06)	1.24 (0.92–1.67)
Routine/manual	2.13 (1.65–2.75)	2.96 (1.22–7.48)	2.12 (1.60–2.83)
Number	1264	129	1035

CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

Table A.5.4 Multivariate models for sickness absence due to IID for cases of all ages and stratified by child and adult age groups

	All ages **0 to 90+ years** OR[a] (95% CI)	**Children** **≥5 to <16 years** OR[a] (95% CI)	**Adults** **≥16 to (<60 women, <65 men) years** OR[a] (95% CI)
Age (years)	0.99 (0.99–0.99)	1.08 (0.96–1.23)	0.98 (0.97–0.99)
Sex			
Female	reference	reference	reference
Male	0.82 (0.68–0.99)	1.29 (0.65–2.60)	0.96 (0.72–1.27)
Ethnicity			
White	reference	reference	reference
Non-White	1.49 (0.93–2.46)	3.31 (0.85–22.04)	2.54 (1.10–6.90)
NS-SEC			
Managerial/ professional	reference	reference	reference
Intermediate	0.97 (0.76–1.23)	1.29 (0.51–3.59)	1.03 (0.73–1.45)
Routine/manual	1.20 (0.95–1.52)	3.19 (1.13–11.48)	1.48 (1.04–2.12)
Number	1846	175	913

[a] Since the absence outcome was common, the odds ratios should not be interpreted as relative risks
CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

Table A.5.5 Multivariate models for sickness absence due to IID for cases of all ages and stratified by child and adult age groups, including symptom severity as an exposure variable

	All ages 0 to 90+ years OR[a] (95% CI)	Children ≥5 to <16 years OR[a] (95% CI)	Adults ≥16 to (<60 women, <65 men) years OR[a] (95% CI)
Age (years)	0.99 (0.98–0.99)	0.99 (0.84–1.17)	0.99 (0.98–1.01)
Sex			
Female	reference	reference	reference
Male	0.81 (0.63–1.04)	1.05 (0.43–2.52)	0.88 (0.62–1.24)
Ethnicity			
White	reference	reference	reference
Non-White	1.33 (0.70–2.61)	2.19 (0.44–16.49)	1.94 (0.68–6.98)
NS-SEC			
Managerial/ professional	reference	reference	reference
Intermediate	0.86 (0.63–1.18)	1.19 (0.36–4.38)	1.01 (0.66–1.56)
Routine/manual	1.10 (0.80–1.50)	4.20 (0.99–29.21)	1.24 (0.81–1.90)
Symptom severity			
Mild	reference	reference	reference
Moderate	3.41 (2.56–4.56)	4.19 (1.62–11.79)	3.73 (2.56–5.49)
Severe	5.64 (4.15–7.72)	8.85 (2.56–41.76)	5.04 (3.30–7.81)
Number	1250	127	695

[a] Since the absence outcome was common, the odds ratios should not be interpreted as relative risks
CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

A.5.5 BOUNDARIES OF SYMPTOM SEVERITY CATEGORIES

The IID symptom severity score ranged from 2 to 40, with a median of 12 and standard deviation of 6.63. In the main analysis, the symptom severity score was converted into tertiles: mild (severity score 2–9), moderate (severity score 10–15) and severe (severity score 16–40). To investigate the robustness of the results, the boundaries of the mild, moderate and severe categories were changed to assess whether this had an impact on the results. The boundaries were changed so that there was a 12 point severity score difference within each category, i.e. mild (severity score 2–14), moderate (severity score 15–27) and severe (severity score 28–40). The results for the most parsimonious model from the main analysis

are displayed in Table A.5.6. The results were similar to those reported in the main analysis, and the estimate for NS-SEC intermediate compared to managerial/professional occupations reached statistical significance. However, there were a small number of cases in the severe category using the altered boundaries (mild [n=875], moderate [n=486] and severe [n=34]).

Table A.5.6 Multivariate model for severe IID symptoms versus mild or moderate symptoms combined for cases ≥5 years of age, with symptom severity category boundaries changed

	OR (95% CI)
Age group (years)	
5–14	reference
15–24	2.63 (1.42–4.93)
25–44	0.89 (0.58–1.36)
45–64	0.65 (0.43–0.99)
65+	0.56 (0.36–0.87)
Sex	
Female	reference
Male	0.91 (0.71–1.17)
Ethnicity	
White	reference
Non-White	1.68 (0.90–3.14)
NS-SEC	
Managerial/professional	reference
Intermediate	1.44 (1.05–1.98)
Routine/manual	2.07 (1.54–2.78)

Based on 1164 cases with complete data, model adjusted for all variables in the table
AIC = 1694.6
CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

A.5.6 LINEAR REGRESSION FOR SYMPTOM SEVERITY SCORE

In the main analysis, the symptom severity score outcome was converted into categories, however converting a continuous variable into categories can result in loss of information. Sensitivity analysis was performed to investigate whether this had an impact on the results. Since the symptom severity score had a positive skew, a log transformation was used to obtain a more normally distributed dependent variable for linear regression analysis. The results for the most parsimonious model from the main analysis are displayed in Table A.5.7.

The results from the linear regression model were similar to the ordinal logistic regression model used in the main analysis, except that male sex was statistically significantly negatively associated with the log of the symptom severity score, and the association between ethnicity and log symptom severity did not reach statistical significance.

Table A.5.7 Multivariate linear regression model for log symptom severity for cases ≥5 years of age

	Multivariate estimate for log IID symptom severity	Std. error
Age group (years)		
5–14	reference	reference
15–24	0.24001*	0.09868
25–44	-0.01739	0.06742
45–64	-0.17201**	0.06450
65+	-0.24346***	0.06856
Sex		
Female	reference	reference
Male	-0.09120*	0.03822
Ethnicity		
White	reference	reference
Non-White	0.17501	0.09863
NS-SEC		
Managerial/professional	reference	reference
Intermediate	0.06192	0.04847
Routine/manual	0.25713***	0.04663

*p <0.05; **p <0.01; ***p <0.001
Based on 1164 cases with complete data, model adjusted for all variables in the table
Adjusted $R^2 = 0.0642$
IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; Std. error = standard error

A.5.7 Analysis with recurrent episodes of IID

In the main analysis, recurrent episodes of IID were removed, thus if a participant experienced more than one episode of IID during follow-up, only information related to the first episode was retained to create a sample of independent observations. Mixed effects models were used to analyse the data with the recurrent episodes included, whilst accounting for clustering at the individual level (since observations from the same individual were considered non-independent). Each participant was given a unique study number. Those with recurrent episodes could be identified by their study number and so this variable was added

as a source of random variation to the model, or the random effect term. The most parsimonious models for the symptom severity and sickness absence outcomes identified in the main analysis were used.

For the symptom severity outcome, a cumulative link mixed model (mixed effects ordinal logistic regression model) (Christensen, 2015), fitted with the adaptive Gauss-Hermite quadrature approximation with 20 quadrature points was used. Maximum likelihood estimates of the parameters were provided using the adaptive Gaussian-Hermite quadrature method, which provides a more accurate approximation of the likelihood compared to the Laplace approximation (Christensen, 2015). The approximation of the likelihood becomes more accurate as more quadrature points are used (UCLA: Statistical Consulting Group, no date-c), in this case 20 were used. Table A.5.8 shows the results of the multivariate mixed effects ordinal logistic regression model, predicting IID symptom severity with repeat episodes from the same individual included. The confidence intervals for some estimates were fairly wide, likely due to the impact of including the random effect term on the power of the analysis (Johnson et al., 2015).

Table A.5.8 Multivariate mixed effects ordinal logistic regression model for severe IID symptoms versus mild or moderate symptoms combined for cases ≥5 years of age

	OR (95% CI)
Age group (years)	
5–14	reference
15–24	9.76 (2.21–43.11)
25–44	0.97 (0.41–2.30)
45–64	0.41 (0.17–0.96)
65+	0.31 (0.12–0.79)
Sex	
Female	reference
Male	0.79 (0.48–1.30)
Ethnicity	
White	reference
Non-White	5.53 (1.36–22.57)
NS-SEC	
Managerial/professional	reference
Intermediate	1.51 (0.80–2.84)
Routine/manual	5.81 (2.67–12.64)
Random effects: Study number, number of groups = 1191	

Based on 1274 cases with complete data, model adjusted for all variables in the table
AIC = 2684.28
CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

Mixed effects logistic regression was used to model the binary sickness absence outcome, with repeat episodes from the same individuals included. A mixed effects model fitted with the adaptive Gauss-Hermite quadrature approximation with 20 quadrature points was used. Table A.5.9 shows the results of the nested multivariate mixed effects models; Model 1 without the symptom severity exposure variable, Model 2 with the symptom severity exposure variable included.

Table A.5.9 Multivariate mixed effects logistic regression models for sickness absence due to IID for cases of school/working age

	Model 1 OR[a] (95% CI)	**Model 2** OR[a] (95% CI)
Age (years)	0.96 (0.93–0.99)	0.98 (0.96–1.00)
Sex		
Female	reference	reference
Male	0.76 (0.39–1.50)	0.77 (0.44–1.35)
Ethnicity		
White	reference	reference
Non-White	8.06 (0.93–69.58)	2.71 (0.54–13.63)
NS-SEC		
Managerial/professional	reference	reference
Intermediate	1.23 (0.51–2.96)	0.99 (0.48–2.05)
Routine/manual	3.73 (1.27–10.93)	1.52 (0.72–3.21)
Symptom severity		
Mild	-	reference
Moderate	-	11.68 (3.34–40.83)
Severe	-	32.69 (5.92–180.64)
Random effects: Study number, number of groups = 837		
Likelihood ratio test Model 1 versus Model 2: χ^2 statistic = 110.51; p <0.001 ***		

[a] Since the absence outcome was common, the odds ratios should not be interpreted as relative risks
Based on 892 cases with complete data
Models adjusted for all variables in the table (Model 1 not adjusted for symptom severity)
AIC: Model1 = 1147.1; Model2 = 1040.6
CI = confidence interval; IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

A.5.8 MULTIPLE IMPUTATION: SYMPTOM SEVERITY

There was a large amount of missing data in the sample of cases aged ≥5 years, used in the main analysis to investigate the symptom severity outcome. The primary exposure of interest NS-SEC had 16% missing data, and the symptom severity outcome had 27% missing data. The variables age, sex and ethnicity were complete and had no missing data. The characteristics of cases with missing data within the NS-SEC and symptom severity variables are shown in Table A.5.10. As can be seen, for the NS-SEC and symptom severity variables, the characteristics of cases with missing data compared to those without missing data were largely similar. Nonetheless, due to the large amount of missing data, multiple imputation was performed to investigate whether the use of listwise deletion in the main analysis had affected the results.

Table A.5.10 Characteristics of cases with missing data within NS-SEC and symptom severity variables

	NS-SEC		Symptom severity	
	Cases without missing data	Cases with missing data	Cases without missing data	Cases with missing data
Number	1616	299	1395	520
Age (years) (mean [SD])	50.0 (21.2)	52.6 (21.3)	47.9 (20.9)	50.8 (22.1)
Male	621 (38.4)	104 (34.8)	536 (38.4)	189 (36.3)
Ethnicity Non-White	60 (3.7)	17 (5.7)	56 (4.0)	21 (4.0)
NS-SEC				
Managerial/professional	-	-	694 (59.4)	255 (57.0)
Intermediate	-	-	224 (19.2)	106 (23.7)
Routine/manual	-	-	251 (21.5)	86 (19.2)
Symptom severity				
Mild	394 (33.7)	76 (33.6)	-	-
Moderate	399 (34.1)	78 (34.5)	-	-
Severe	376 (32.2)	72 (31.9)	-	-

Figures expressed as number (%) except where stated otherwise
NS-SEC = National Statistics Socioeconomic Classification; SD = standard deviation

Nature of the missing data

Multiple imputation can be performed in situations where missing data is either MCAR or MAR. To explore this, univariate logistic regression analyses were performed to investigate whether the missingness in the symptom severity and NS-SEC variables could have been

explained by other variables in the dataset. For the dependent variables 'Not missing' was coded as 0, and 'Missing' as 1. The results are displayed in Table A.5.11.

Table A.5.11 Univariate logistic regression for cases with missing data versus cases without missing data

	Univariate logistic regression for missing data vs. not missing OR (95% CI)	
	Symptom severity	**NS-SEC**
Age (years)	1.01 (1.00–1.01)	1.01 (1.00–1.02)
Male[a]	0.92 (0.74–1.13)	0.85 (0.66–1.10)
Ethnicity Non White[b]	1.01 (0.59–1.65)	1.56 (0.87–2.66)
IMD quintile[c]		
4	1.02 (0.77–1.34)	1.99 (1.38–2.91)
3	1.08 (0.81–1.45)	1.48 (1.00–2.22)
2	1.40 (0.97–2.03)	2.20 (1.36–3.53)
1 (most deprived)	1.23 (0.80–1.86)	2.89 (1.75–4.74)
Rural residency[d]	1.11 (0.89–1.38)	1.25 (0.96–1.62)
Study type Cohort[e]	2.42 (1.96–2.99)	0.89 (0.70–1.15)
Absent[f]	1.15 (0.93–1.42)	0.71 (0.55–0.92)
Absence duration (days)	0.92 (0.84–0.99)	1.07 (0.98–1.16)
NS-SEC Intermediate[g]	1.29 (0.98–1.69)	-
NS-SEC Routine/manual[g]	0.93 (0.70–1.23)	-
Symptom severity score	-	1.00 (0.98–1.02)
Travelled abroad[h]	0.78 (0.56–1.08)	0.89 (0.59–1.30)

[a] reference category = Female
[b] reference category = Ethnicity White
[c] reference category = IMD quintile 5 (least deprived)
[d] reference category = Urban residency
[e] reference category = Study type GP Presentation
[f] reference category = Not absent
[g] reference category = NS-SEC Managerial/professional occupations
[h] reference category = No foreign travel before illness
CI = confidence interval; GP = general practice; IMD = Index of Multiple Deprivation; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

As can be seen, certain variables within the dataset were related to the missingness within the symptom severity and NS-SEC variables. Missingness within the symptom severity score variable was statistically significantly related to age, study type and the duration of absence among absentees. Older cases and those recruited to the Cohort study as opposed to the GP Presentation study were more likely to have missing values within the symptom severity score variable. Absentees with longer durations of absence were less likely to have missing data within the symptom severity score variable.

Missingness within the NS-SEC variable was statistically significantly related to age, the IMD quintiles and absence due to IID. Those in the most deprived IMD quintile (1), second most deprived (2) and those in the second least deprived quintile (4) all had greater odds of having missing values within the NS-SEC variable, compared to the least deprived quintile (5). Younger cases and those who were absent were less likely to have missing data within the NS-SEC variable. The significant associations observed give support to the missing mechanism being MAR.

Variables included in multiple imputation model

MICE was used to impute the missing data. Table A.5.12 shows the variables used in the imputation model, grouped by the main reason for which they were chosen. The variables selected to be used in the imputation model had varying distributions and therefore the imputation method for each variable was specified individually (Table A.5.12).

Table A.5.12 Predictor variables used in multiple imputation model

Variable Name	Class	Levels	Description	MI Method
Variables used in final model				
initial_age	integer	-	age in years	†
sex	factor	2	female, male	†
ethnic_group	factor	2	White, Non-White	†
severity_score	integer	-	symptom severity score	pmm
nssec3	ordinal factor	3	managerial/professional, intermediate, routine/manual	polr
Variables related to missingness in NS-SEC				
imdquintile	ordinal factor	5	Index of Multiple Deprivation quintile	polr
absence	factor	2	absent work/school/daily activities: yes, no	logreg
Variables related to missingness in symptom severity				
study	factor	2	GP Presentation study, Cohort study	†
absence_days	integer	-	how many days absent	pmm
Variables related to symptom severity				
diarrhoea	factor	2	had diarrhoea: yes, no	logreg
diarrhoea_days	integer	-	how many days diarrhoea	pmm
diarrhoea_blood	factor	2	had diarrhoea with blood: yes, no	logreg
diarrhoea_blood_days	integer	-	how many days diarrhoea with blood	pmm
vomiting	factor	2	had vomiting: yes, no	logreg
vomiting_days	integer	-	how many days vomiting	pmm
nausea	factor	2	had nausea: yes, no	logreg
nausea_days	integer	-	how many days nausea	pmm
abdominal_pain	factor	2	had abdominal pain: yes, no	logreg

loss_of_appetite	factor	2	had loss of appetite: yes, no	logreg
fever	factor	2	had fever: yes, no	logreg
headache	factor	2	had headache: yes, no	logreg
cough_nose_throat	factor	2	had cough/runny nose/sore throat: yes, no	logreg
Variables related to NS-SEC				
occupation	factor	8	occupation of main earner	polyreg
employment_status	factor	7	employment status of main earner	polyreg
employment_type	factor	3	employee, self-employed with employees, self-employed no employees	polyreg
employees	factor	2	number of employees: 0-24, >=25	logreg
supervisorstatus	factor	2	supervise others: yes, no	logreg
urind_cat	factor	2	urban, rural residency	logreg
travel	factor	2	foreign travel before illness: yes, no	logreg

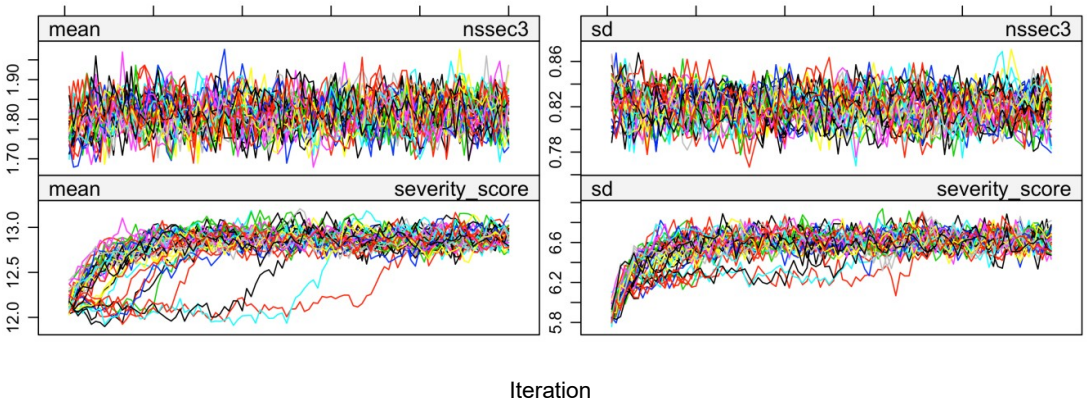
Logreg = logistic regression; pmm = predictive mean matching; polr = ordinal logistic regression; polyreg = multinomial regression; † no method used as no missing data
MI = multiple imputation; NS-SEC = National Statistics Socioeconomic Classification

Running MICE and assessing convergence

For this analysis, 43.6% of cases had missing data on one or more variables, and therefore approximately 40 multiply imputed datasets were required. Due to the large amount of missing data 100 iterations were used.

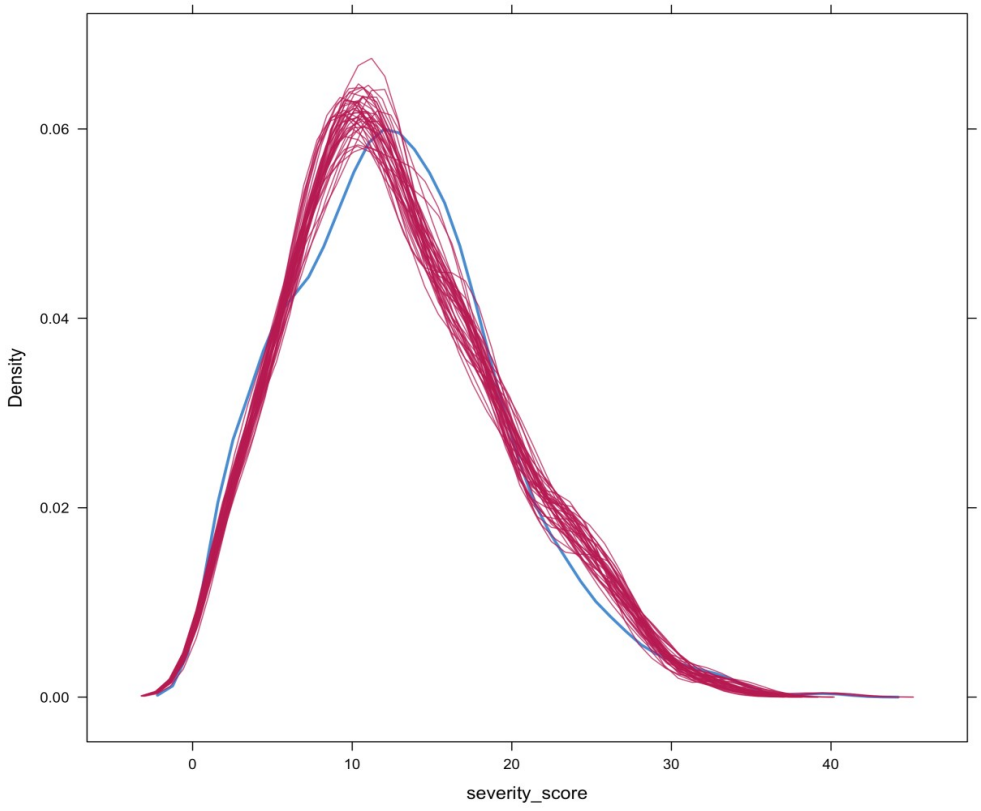
After running the multiple imputation model, convergence was checked by plotting the mean and variance of the imputed values per iteration for each variable (Figure A.5.4). For the NS-SEC variable, the coloured lines freely intermingled with each other without showing any trends, indicating healthy convergence. For the symptom severity score variable, there was a strong initial trend due to the randomly drawn initial value being too low for the severity score variable. By 40 iterations this initial trend was corrected and within 80 iterations all of the lines were intermingling well. The symptom severity score variable needed a greater number of iterations to converge and become stable, compared to the NS-SEC variable, possibly because of the large amount of missing data within the symptom severity score variable (27% missing).

Figure A.5.4 Means and standard deviations per iteration of the imputed values of NS-SEC and symptom severity score



Additionally to check the plausibility of the imputed values, the densities for the observed and imputed values were plotted for the symptom severity score variable (Figure A.5.5), with the observed values in blue and imputed values in pink. As can be seen the imputed values were very similar to the observed values and the positive skew of the data was preserved.

Figure A.5.5 Density plot showing densities for the observed and imputed values for the symptom severity score



Analysis with multiply imputed datasets

In the main analysis a multivariate model was created to investigate the effect of NS-SEC on IID symptom severity whilst controlling for age, sex and ethnicity using listwise deletion to handle the missing data. The results of this analysis are displayed in Table A.5.13 (Model 1). In addition, the results for the same model using the multiply imputed datasets are shown, firstly with the deletion of cases that had imputed dependent variable values (MID method – Model 2), and secondly with the imputed dependent variable values retained in the analysis (Model 3).

Table A.5.13 Multivariate analysis for severe IID symptoms versus mild or moderate symptoms combined for cases ≥5 years of age

	Model 1 Listwise deletion[1]		Model 2 MID[2]		Model 3 Imputed dependent variable retained[3]	
	OR	SE	OR	SE	OR	SE
NS-SEC Intermediate [a]	1.21	0.14423	1.17	0.13661	1.26*	0.11728
NS-SEC Routine/manual [a]	2.18***	0.13947	1.88***	0.13473	1.80***	0.11749
Age group 15–24 years [b]	2.70**	0.31034	2.40**	0.27764	2.35***	0.24132
Age group 25–44 years [b]	0.96	0.19769	0.95	0.18533	1.08	0.16394
Age group 45–64 years [b]	0.69*	0.18957	0.67*	0.17791	0.72*	0.15414
Age group 65+ years [b]	0.60*	0.20235	0.63*	0.18670	0.65**	0.16067
Male [c]	0.90	0.11340	0.87	0.10379	0.90	0.09119
Ethnicity Non-White [d]	2.03*	0.30005	1.45	0.26107	1.26	0.22838

*p <0.05; **p <0.01; ***p <0.001
[a] reference category = NS-SEC Managerial/professional occupations
[b] reference category = Age group 5–14 years
[c] reference category = Female
[d] reference category = Ethnicity White
[1] Based on 1164 cases, model adjusted for all variables in the table
[2] Based on 1395 cases, model adjusted for all variables in the table
[3] Based on 1915 cases, model adjusted for all variables in the table
IID = infectious intestinal disease; MID = multiple imputation then deletion; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio; SE = standard error

It has been suggested that the MID method can provide less variable point estimates and more accurate standard error estimates compared to analyses that retain the imputed dependent variable values, particularly when there are large amounts of missing data within the dependent variable (von Hippel, 2007). For this analysis, Model 3 with the imputed dependent variable values retained had the smallest standard error estimates compared to the

other models. The results across all of the models were largely similar, although the magnitude of the association between NS-SEC and symptom severity was slightly weaker when the multiply imputed datasets were used compared to listwise deletion. Ethnicity was not associated with symptom severity in models that used the multiply imputed datasets.

A.5.9 MULTIPLE IMPUTATION: SICKNESS ABSENCE

In the main analysis the sickness absence outcome was investigated using cases of school or working age. Within this dataset, the primary exposure of interest NS-SEC had 13% missing data, symptom severity had 24% missing data and the absence outcome had 2.7% missing data. The variables age, sex and ethnicity were complete and had no missing data. The characteristics of cases with missing data within the NS-SEC and sickness absence variables are shown in Table A.5.14. As can be seen, the characteristics of cases with missing data compared to those without missing data were largely similar. Proportionately fewer cases with missing NS-SEC reported sickness absence compared to cases without missing NS-SEC, and a larger proportion of cases with missing sickness absence were of Non-White ethnicity compared to cases who answered the absence question.

Table A.5.14 Characteristics of cases with missing data within NS-SEC and sickness absence variables

	NS-SEC		**Sickness absence**	
	Cases without missing data	Cases with missing data	Cases without missing data	Cases with missing data
Number	1105	165	1236	34
Age (years) (mean [SD])	37.6 (17.2)	37.9 (17.2)	37.8 (17.0)	34.3 (20.8)
Male	456 (41.3)	65 (39.4)	502 (40.6)	19 (55.9)
Ethnicity Non-White	52 (4.7)	13 (7.9)	59 (4.8)	6 (17.6)
NS-SEC				
Managerial/professional	-	-	651 (59.8)	11 (64.7)
Intermediate	-	-	212 (19.5)	3 (17.6)
Routine/manual	-	-	225 (20.7)	3 (17.6)
Symptom severity				
Mild	257 (31.1)	43 (31.2)	294 (31.1)	6 (30.0)
Moderate	288 (34.8)	46 (33.3)	326 (34.5)	8 (40.0)
Severe	282 (34.1)	49 (35.5)	325 (34.4)	6 (30.0)
Absent	694 (63.8)	81 (54.7)	-	-

Figures expressed as number (%) except where stated otherwise
NS-SEC = National Statistics Socioeconomic Classification; SD = standard deviation

Nature of missing data

To explore whether the missing data were MAR or MCAR, univariate logistic regression analyses were performed to investigate whether the missingness in the absence, symptom severity and NS-SEC variables could have been explained by other variables in the dataset. For the dependent variables 'Not missing' was coded as 0, and 'Missing' as 1. The results are displayed in Table A.5.15.

Table A.5.15 Univariate logistic regression for cases with missing data versus cases without missing data

	Univariate logistic regression for missing data vs. not missing OR (95% CI)		
	Symptom severity	**NS-SEC**	**Absence**
Age (years)	1.00 (0.99–1.00)	1.00 (0.99–1.01)	0.99 (0.97–1.01)
Male[a]	1.19 (0.92–1.54)	0.93 (0.66–1.29)	1.85 (0.93–3.74)
Ethnicity Non White[b]	1.22 (0.68–2.11)	1.73 (0.89–3.16)	4.27 (1.55–10.08)
IMD quintile[c]			
4	1.16 (0.81–1.67)	2.24 (1.34–3.90)	0.23 (0.05–0.78)
3	1.12 (0.76–1.65)	1.67 (0.95–3.01)	1.01 (0.40–2.59)
2	1.69 (1.06–2.68)	2.91 (1.55–5.54)	0.95 (0.25–2.98)
1 (most deprived)	1.62 (0.98–2.65)	4.64 (2.50–8.78)	2.47 (0.90–6.62)
Rural residency[d]	1.12 (0.83–1.47)	0.88 (0.60–1.27)	0.67 (0.27–1.48)
Study type Cohort[e]	2.33 (1.78–3.06)	0.91 (0.66–1.26)	1.05 (0.53–2.10)
Absent[f]	1.30 (0.99–1.73)	0.69 (0.49–0.97)	-
Absence duration (days)	0.91 (0.83–1.00)	1.07 (0.96–1.17)	-
NS-SEC Intermediate[g]	1.32 (0.93–1.84)	-	0.84 (0.19–2.71)
NS-SEC Routine/manual[g]	0.83 (0.57–1.19)	-	0.79 (0.18–2.55)
Symptom severity score	-	1.01 (0.98–1.04)	0.97 (0.90–1.04)
Travelled abroad[h]	0.84 (0.55–1.23)	0.65 (0.36–1.10)	1.57 (0.58–3.64)

[a] reference category = Female
[b] reference category = Ethnicity White
[c] reference category = IMD quintile 5 (least deprived)
[d] reference category = Urban residency
[e] reference category = Study type GP Presentation
[f] reference category = Not absent
[g] reference category = NS-SEC Managerial/professional occupations
[h] reference category = No foreign travel before illness
CI = confidence interval; GP = general practice; IMD = Index of Multiple Deprivation; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio

As can be seen in Table A.5.15, certain variables within the dataset were related to the missingness within the absence, symptom severity and NS-SEC variables. Missingness within the absence variable was statistically significantly related to ethnicity and one quintile

of the IMD. Those of Non-White ethnicity compared to White were more likely to have missing values within the absence variable.

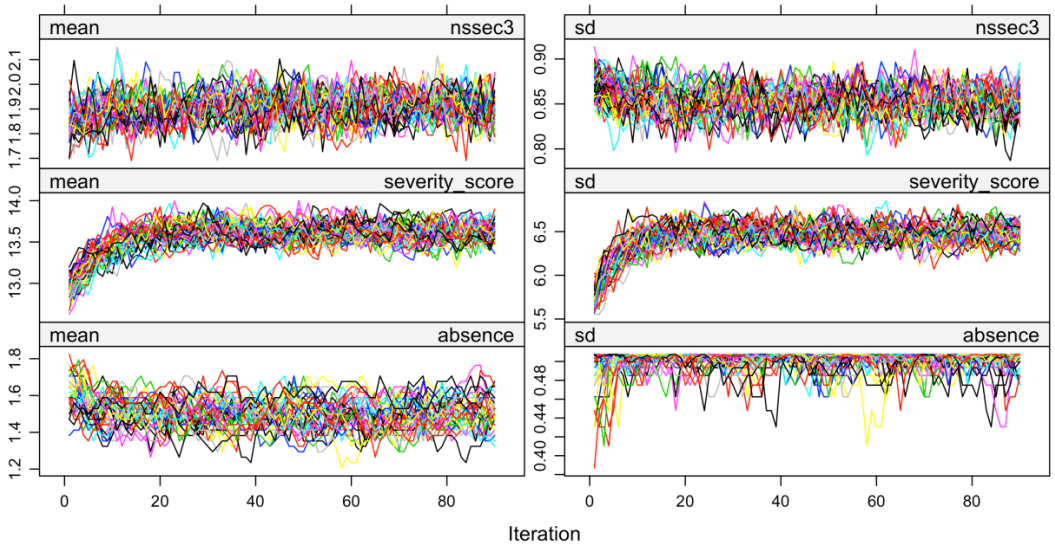
Missingness within the NS-SEC variable was statistically significantly related to the IMD quintiles and absence due to IID, as observed previously in the dataset containing cases ≥5 years of age. Missingness within the symptom severity score variable was statistically significantly related to study type, the duration of absence among absentees (as observed previously) and one quintile of the IMD. Again, the significant associations observed give support to the missing mechanism being MAR.

Running MICE and assessing convergence

The variables included in the imputation model were the same as those for the previous multiple imputation using the dataset containing cases ≥5 years of age, since the variables of interest to be imputed were the same with the addition of the absence variable. Again, 40 multiply imputed datasets were used, since 38.8% of cases had missing data within one or more variables. Since convergence was achieved by 80 iterations in the previous multiple imputation model, 90 iterations were specified for this analysis to save computational time.

Convergence was checked by plotting the mean and variance of the imputed values per iteration (Figure A.5.6). For the sickness absence, NS-SEC and symptom severity variables, the lines freely intermingled with each other without showing any trends, indicating that convergence was achieved.

Figure A.5.6 Means and standard deviations per iteration of the imputed values of NS-SEC, symptom severity and sickness absence



Analysis with multiply imputed datasets

In the main analysis a multivariate model was created to investigate the effect of NS-SEC on sickness absence due to IID whilst controlling for age, sex and ethnicity. A second model was created to investigate the effect of adding symptom severity as a covariate to the first model. Listwise deletion was used to handle the missing data. The results of these analyses are displayed in Tables A.5.16 and A.5.17 (Model 1). In addition, the results for the same models using the multiply imputed datasets are shown, firstly with the deletion of cases that had imputed dependent variable values (MID method – Model 2), and secondly with the imputed dependent variable values retained in the analysis (Model 3).

For the most part, the models with the imputed dependent variable values retained had the smallest standard error estimates compared to the other models. The models based on the imputed datasets had smaller standard error estimates compared to the listwise deletion model used in the main analysis. Despite this, the results across all of the models were largely similar. The magnitude of the association between NS-SEC and sickness absence was weaker when the multiply imputed datasets were used compared to listwise deletion. For the models that did not include the symptom severity covariate, ethnicity was not associated with sickness absence when the multiply imputed datasets were analysed.

Table A.5.16 Multivariate analysis for sickness absence due to IID for cases of school/working age

	Model 1		Model 2		Model 3	
	Listwise deletion[1]		MID[2]		Imputed dependent variable retained[3]	
	OR†	SE	OR†	SE	OR†	SE
NS-SEC Intermediate[a]	1.13	0.19576	1.04	0.16139	1.04	0.15921
NS-SEC Routine/manual[a]	1.83**	0.19388	1.40*	0.16309	1.38*	0.16133
Age (years)	0.98***	0.00451	0.98***	0.00362	0.98***	0.00360
Male[b]	0.91	0.15081	0.98	0.12215	0.99	0.12222
Ethnicity Non-White[c]	2.58*	0.46022	1.85	0.32440	1.78	0.32417

*p <0.05; **p <0.01; ***p <0.001
[a] reference category = NS-SEC Managerial/professional occupations
[b] reference category = Female
[c] reference category = Ethnicity White
[1] Based on 818 cases, model adjusted for all variables in the table
[2] Based on 1236 cases, model adjusted for all variables in the table
[3] Based on 1270 cases, model adjusted for all variables in the table
† Since the absence outcome was common, the odds ratios should not be interpreted as relative risks
MID = multiple imputation then deletion; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio; SE = standard error

Table A.5.17 Multivariate analysis for sickness absence due to IID for cases of school/working age with symptom severity as additional covariate

	Model 1 Listwise deletion[1]		Model 2 MID[2]		Model 3 Imputed dependent variable retained[3]	
	OR†	SE	OR†	SE	OR†	SE
NS-SEC Intermediate[a]	1.05	0.20843	0.95	0.17157	0.96	0.16922
NS-SEC Routine/manual[a]	1.38	0.20674	1.12	0.17184	1.10	0.17041
Age (years)	0.99*	0.00473	0.99***	0.00379	0.99**	0.00378
Male[b]	0.92	0.15976	0.99	0.12939	1.00	0.12994
Ethnicity Non-White[c]	1.91	0.47473	1.67	0.33873	1.63	0.33982
Symptom severity Moderate[d]	3.60***	0.17932	3.29***	0.15177	3.25***	0.15158
Symptom severity Severe[d]	5.27***	0.20556	5.01***	0.16737	5.01***	0.16535

*p <0.05; **p <0.01; ***p <0.001
[a] reference category = NS-SEC Managerial/professional occupations
[b] reference category = Female
[c] reference category = Ethnicity White
[d] reference category = Symptom severity Mild
[1] Based on 818 cases, model adjusted for all variables in the table
[2] Based on 1236 cases, model adjusted for all variables in the table
[3] Based on 1270 cases, model adjusted for all variables in the table
† Since the absence outcome was common, the odds ratios should not be interpreted as relative risks
MID = multiple imputation then deletion; NS-SEC = National Statistics Socioeconomic Classification; OR = odds ratio; SE = standard error

A.5.10 SICKNESS ABSENCE DURATION DUE TO IID

Cases who were absent from work, school or their normal daily activities were asked to report how many days they were absent for in total. A final analysis was performed using absentee IID cases of school or working age, to investigate predictors of sickness absence duration due to IID.

Descriptive statistics

From the original sample of cases of school or working age (n=1270), 61% were absent from work, school or daily activities due to their illness (n=775). Of those who were absent, 759 cases provided information about the duration of their absence (97.9%). The number of days the absentees were absent ranged from 1–28 days, with a mean and variance of 2.6 and 4.2,

respectively. Characteristics of the absentees of school or working age are displayed in Table A.5.18.

Table A.5.18 Characteristics of absentee IID cases of school or working age

	Cases school/working age (n=775)				
	Percentage within each category of NS-SEC				
	Managerial/ professional (n=400)	Intermediate (n=133)	Routine/ manual (n=161)	p-value[a]	All cases[b] (n=775)
Age (years) mean (SD)	35.7 (17.5)	36.7 (17.5)	36.4 (17.2)	0.832	36 (17.4)
Male	38.5	37.6	50.3	0.024	40.6
Ethnicity Non-White	5.2	8.3	6.2	0.447	5.9
Rural residence	30.9	27.8	21.7	0.092	27.9
Travelled before illness	13	14.3	7.5	0.122	11.6
Symptom severity					
Mild	39.6	33.7	25.8	0.021	35.8
Moderate	34.1	33.7	32.8		33.0
Severe	26.3	32.6	41.4		31.3
Absence duration (days) mean (SD)	2.3 (1.7)	2.8 (2.8)	2.8 (2.0)	0.009	2.6 (2.1)

IID = infectious intestinal disease; NS-SEC = National Statistics Socioeconomic Classification; SD = standard deviation
Figures expressed as percentages except where stated otherwise
[a] Statistical significance of relationship between NS-SEC and each variable, tested using χ^2 test and one-way analysis of variance (ANOVA) for age and absence duration
[b] Total number of cases includes those with missing NS-SEC
Missing data (%): Urban/rural = 0.3; NS-SEC = 10.5; Symptom severity = 25.3; Absence duration = 2.1

Absentee IID cases in routine/manual compared to managerial/professional occupations were less likely to be female (χ^2 7.4; p-value 0.024). Age, ethnicity, urban/rural residency and foreign travel were not statistically significantly associated with NS-SEC (Table A.5.18). Of the 775 absentee IID cases of school or working age, 501 (65%) had complete data for the variables of interest and were included in the univariate and multivariate analyses.

Univariate analysis

Table A.5.19 shows the incident rate ratios and 95% confidence intervals for the univariate relationships between the independent variables and the absence duration dependent variable, using negative binomial regression. Ethnicity, NS-SEC and symptom severity were statistically significantly associated with the duration of absence due to IID.

Table A.5.19 Univariate negative binomial regression for sickness absence duration due to IID for absentee cases of school/working age

	Absence duration IRR (95% CI) Cases with complete data school/working age (n=501)
Age (years)	1.00 (1.00–1.01)
Sex	
Female	reference
Male	1.10 (0.97–1.25)
Ethnicity	
White	reference
Non-White	1.48 (1.16–1.87)
NS-SEC	
Managerial/professional	reference
Intermediate	1.21 (1.02–1.43)
Routine/manual	1.23 (1.05–1.42)
Residence	
Urban	reference
Rural	0.98 (0.85–1.13)
Travelled before illness	
No	reference
Yes	0.95 (0.78–1.16)
Symptom severity	
Mild	reference
Moderate	1.32 (1.12–1.55)
Severe	2.03 (1.74–2.36)

CI = confidence interval; IID = infectious intestinal disease; IRR = incident rate ratio; NS-SEC = National Statistics Socioeconomic Classification

Multivariate analysis

Hierarchical multivariate negative binomial regression was performed to assess the relationship between SES and absence duration due to IID. Table A.5.20 shows the coefficients and standard errors for six nested models and their summary statistics for comparison. The dependent variable for all models was absence duration in days. Model 1 shows the results of multivariate negative binomial regression with age, sex and ethnicity as the exposures. Model 2 shows the results with age, sex and ethnicity as the exposures, with the addition of NS-SEC as the primary exposure of interest. The addition of NS-SEC statistically significantly improved the model fit when comparing the likelihoods of Model 2 and 1 using the likelihood ratio chi-square statistic (Likelihood ratio χ^2 7.2; p-value 0.027) (Table A.5.21). Absentees in routine/manual compared to managerial/professional occupations experienced longer durations of absence due to IID (IRR 1.21; 95% CI 1.04–1.40) (Table A.5.22). There was no improvement in the model fit when the variables urban/rural residency and recent foreign travel were added to Model 2 (Table A.5.21).

A final model was created to investigate the effect of IID symptom severity on the absence duration outcome. The incident rate ratios and 95% confidence intervals for Models 2 and 6 are presented in Table A.5.22. When symptom severity was added to Model 2, the association between NS-SEC and the absence duration outcome was attenuated and rendered non-significant (IRR 1.08; 95% CI 0.94–1.25). Symptom severity was positively associated with absence duration, and the incident rate ratio for absence duration increased as symptom severity increased (Table A.5.22). Across all models, there was a statistically significant relationship between ethnicity and absence duration (Table A.5.20). Those of Non-White ethnicity compared to White were more likely to be absent for longer due to IID, however it must be noted that the number of participants in the Non-White ethnic group was very small (n=30).

Table A.5.20 Nested multivariate negative binomial regression models for sickness absence duration due to IID for cases of school/working age

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Age (years)	0.00263	0.00238	0.00236	0.00246	0.00244	0.00436*
std. error	0.00189	0.00188	0.00189	0.00189	0.0019	0.00182
Male [a]	0.08398	0.06712	0.06676	0.06698	0.06672	0.09731
std. error	0.06535	0.06546	0.06558	0.06544	0.06556	0.06215
Ethnicity Non-White [b]	0.40556***	0.38428**	0.38580**	0.38642**	0.38749**	0.32724**
std. error	0.12243	0.12204	0.12284	0.1222	0.12297	0.11487
NS-SEC Intermediate [c]		0.15351	0.15354	0.15321	0.15324	0.10292
std. error		0.08579	0.08579	0.08579	0.08578	0.08148
NS-SEC Routine/manual[c]		0.18797*	0.18857*	0.18489*	0.18535*	0.08149
std. error		0.07595	0.07631	0.07621	0.07661	0.07269
Rural residency [d]			0.00702		0.00504	
std. error			0.07437		0.07445	
Foreign travel [e]				-0.04735	-0.04697	
std. error				0.09953	0.09966	
Symptoms Moderate [f]						0.29591***
std. error						0.08158
Symptoms Severe [f]						0.70612***
std. error						0.07794
Log-likelihood	-957.5	-953.9	-953.9	-953.8	-953.8	-912.7
Deviance	429.7	427.7	427.7	427.8	427.8	400.9
AIC	1925	1921.8	1923.8	1923.5	1925.5	1843.3
BIC	1946.1	1951.3	1957.5	1957.3	1963.5	1881.3
Number	501	501	501	501	501	501

*p <0.05; **p <0.01; ***p <0.001
[a] reference category = Female
[b] reference category = Ethnicity White
[c] reference category = NS-SEC Managerial/professional occupations
[d] reference category = Urban residency
[e] reference category = No foreign travel
[f] reference category = Symptom severity Mild
AIC = Akaike information criterion; BIC = Bayesian information criterion; NS-SEC = National Statistics Socioeconomic Classification; std. error = standard error

Table A.5.21 Likelihood ratio tests for comparison of nested models

Test	Likelihood ratio χ^2 statistic	p-value
Model 1 versus Model 2	7.2426	0.02675*
Model 2 versus Model 3	0.0089	0.9248
Model 2 versus Model 4	0.2275	0.6334
Model 2 versus Model 5	0.2321	0.8904
Model 2 versus Model 6	82.451	<0.001***

*p <0.05; **p <0.01; ***p <0.001

Table A.5.22 Multivariate Models 2 and 6 for sickness absence duration due to IID for cases of school/working age

	Model 2	Model 6
	IRR (95% CI)	IRR (95% CI)
Age (years)	1.00 (1.00–1.01)	1.00 (1.00–1.01)
Sex		
Female	reference	reference
Male	1.07 (0.94–1.22)	1.10 (0.98–1.24)
Ethnicity		
White	reference	reference
Non-White	1.47 (1.15–1.86)	1.39 (1.10–1.73)
NS-SEC		
Managerial/professional	reference	reference
Intermediate	1.17 (0.98–1.38)	1.11 (0.94–1.30)
Routine/manual	1.21 (1.04–1.40)	1.08 (0.94–1.25)
Symptom severity		
Mild		reference
Moderate		1.34 (1.15–1.58)
Severe		2.03 (1.74–2.36)

CI = confidence interval; IID = infectious intestinal disease; IRR = incident rate ratio; NS-SEC = National Statistics Socioeconomic Classification
Cases with complete data school/working age (n=501)

Appendices to Chapter 6

The appendix for Chapter 6 features exploratory analyses, investigations of model assumptions and sensitivity analyses that were conducted to assess the robustness of the main results from the analysis of HES data presented in Chapter 6.

A.6.1 Exploratory analysis using GAMs

GAMs were used to visually assess the relationships between the continuous independent variables and the scaled risk of the hospitalisation outcomes.

For adults, the GAMs suggested approximately linear relationships between income deprivation and the scaled rate of emergency hospital admission for IID (Figures A.6.2 and A.6.3). For children, however the relationship was rather more curve-linear (Figure A.6.1). Additionally, the relationships between the scaled rate of emergency hospital admissions and ethnicity (Figures A.6.4 to A.6.6) and long-term health problems for adults aged 65+ years (Figure A.6.9) appeared non-linear. It was therefore decided that these variables (income deprivation, ethnicity and long-term health problems) would be included as categorical variables when modeling the emergency hospital admission rate outcome. Additional GAMS showed that the distance to health service variables were approximately linearly related to the emergency hospital admission rate outcome (data not shown), therefore these variables were retained as continuous variables.

Income deprivation was categorised as quintiles, since the IMD (of which income deprivation is a domain) is commonly operationalised as quintiles. The average percentage of those of White ethnicity was relatively similar for children, adults and older adults, and therefore to aid interpretation the ethnicity variable was categorised with the boundaries: $\leq$70%; >70% to $\leq$90%; >90% for all age groups. In contrast, the average percentage of those with long-term health problems varied markedly across the age groups, and therefore it was more appropriate to categorise this variable as quartiles for each separate age group. Similar results to those presented in the main analysis were observed when the ethnicity variable was operationalised as tertiles, and when the long-term health problem variable was operationalised as quintiles (data not shown).

Figures A.6.1 to A.6.3 GAMs showing the shape of the relationship between income deprivation and IID emergency hospital admission rates, adjusted for ethnicity, long-term health problems and geographical variables

The Y axis in the plots shows the scaled rate of emergency hospital admission for IID

Figure A.6.1 Children aged 0–14 years

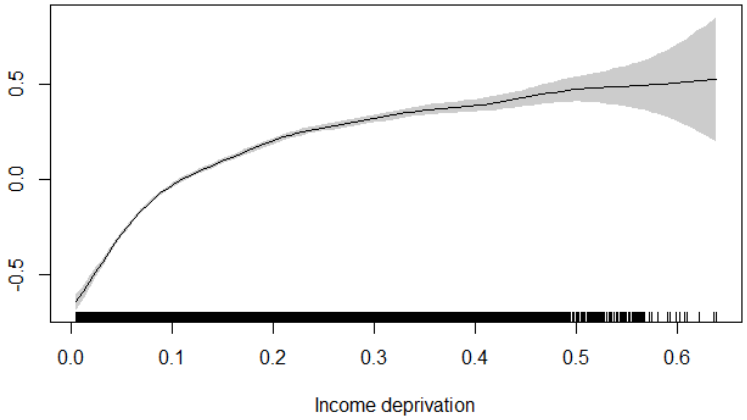


Figure A.6.2 Adults aged 15–64 years

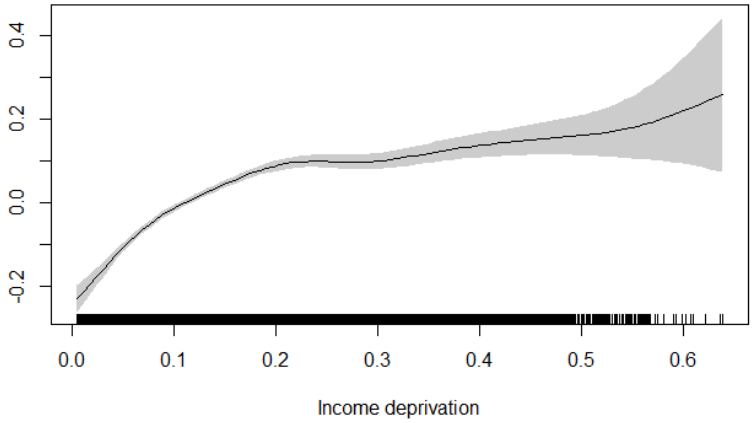
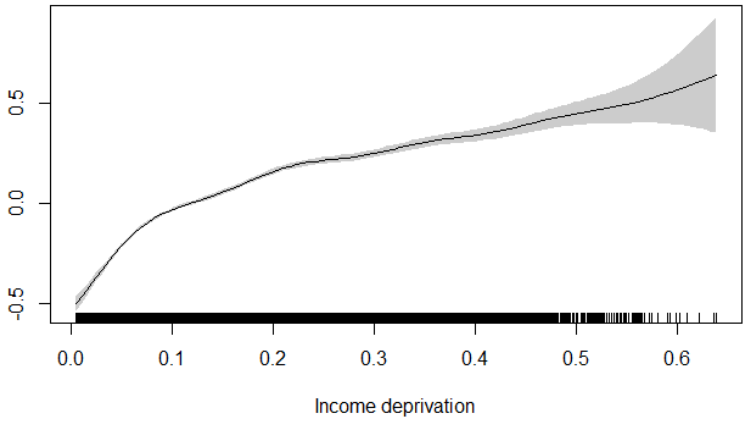


Figure A.6.3 Adults aged 65+ years



Figures A.6.4 to A.6.6 GAMs showing the shape of the relationship between the proportion of the population of White ethnicity and IID emergency hospital admission rates, adjusted for income deprivation, long-term health problems and geographical variables

The Y axis in the plots shows the scaled rate of emergency hospital admission for IID

Figure A.6.4 Children aged 0–14 years

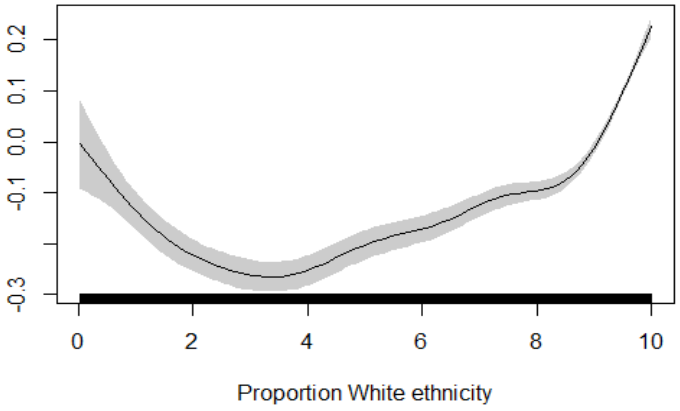


Figure A.6.5 Adults aged 15–64 years

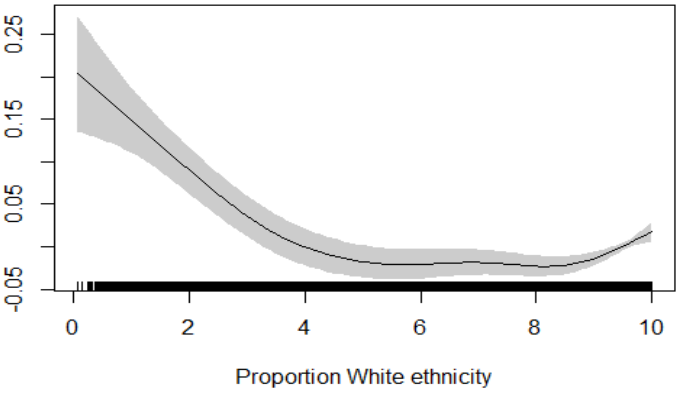
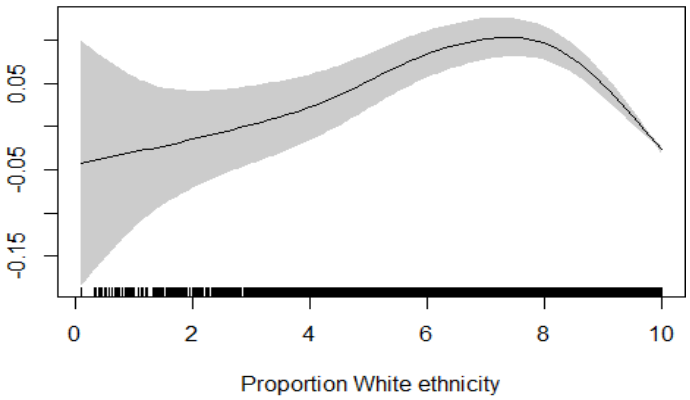


Figure A.6.6 Adults aged 65+ years



Figures A.6.7 to A.6.9 GAMs showing the shape of the relationship between the proportion of the population with a long-term health problem and IID emergency hospital admission rates, adjusted for income deprivation, ethnicity and geographical variables

The Y axis in the plots shows the scaled rate of emergency hospital admission for IID

Figure A.6.7 Children aged 0–14 years

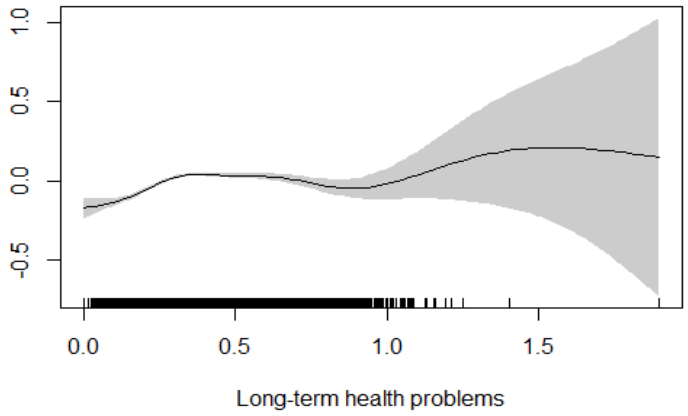


Figure A.6.8 Adults aged 15–64 years

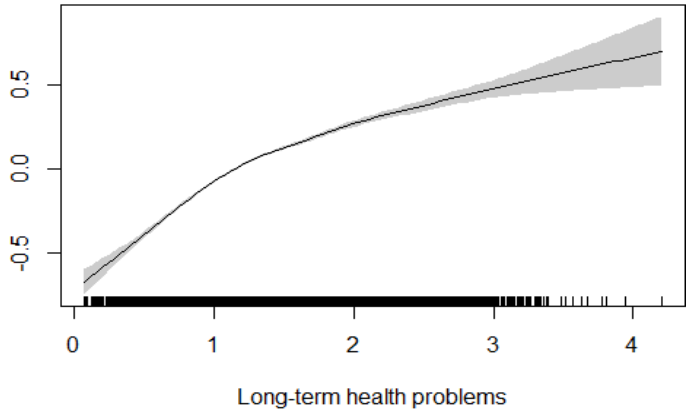
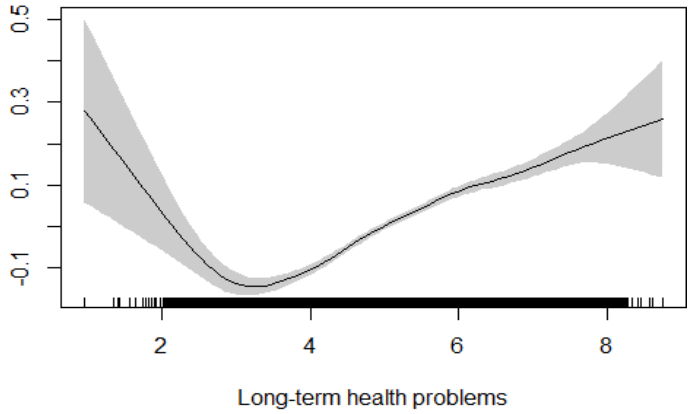


Figure A.6.9 Adults aged 65+ years



Approximate linear relationships were observed between income deprivation and the scaled risk of log admission days per admission for IID (Figures A.6.10 to A.6.12). IID admission duration tended to increase with increasing income deprivation for children and adults aged 65+ years. For adults aged 15–64 years, the scaled risk of log admission days per admission did not appear to increase with increasing income deprivation in a dose-response fashion, when ethnicity, long-term health problems and the geographical variables were accounted for (Figure A.6.11). This finding was also observed in the main results presented in Chapter 6. Additional GAMs showed that all of the continuous variables were approximately linearly related to the admission duration outcome (data not shown), and therefore these variables were retained as continuous variables when modeling this outcome.

Figures A.6.10 to A.6.12 GAMs showing the shape of the relationships between income deprivation and log IID admission days per admission, adjusted for ethnicity, long-term health problems and geographical variables

The Y axis in the plots shows the scaled risk of log admission duration for IID

Figure A.6.10 Children aged 0–14 years

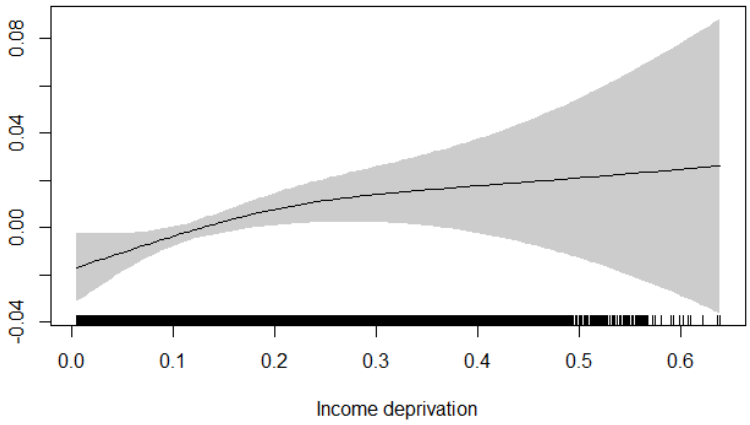


Figure A.6.11 Adults aged 15–64 years

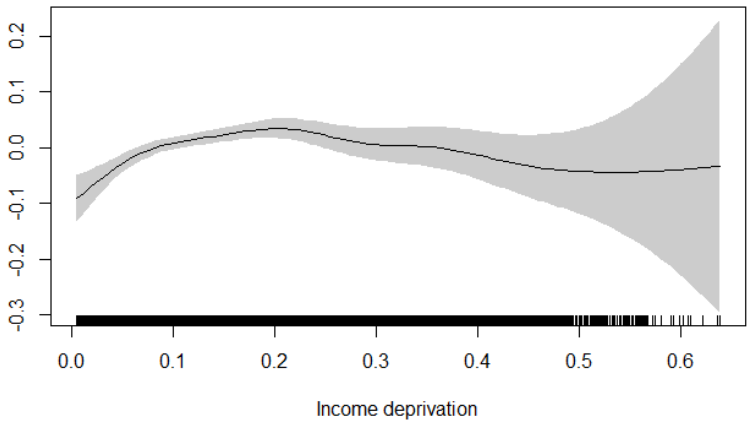
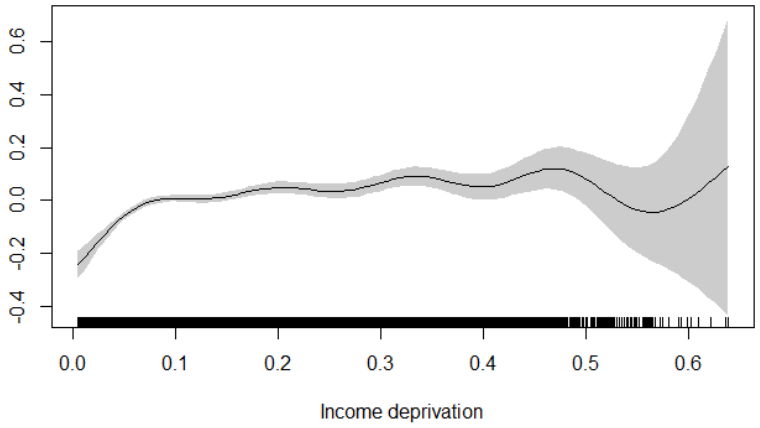


Figure A.6.12 Adults aged 65+ years



A.6.2 LINEAR REGRESSION MODEL ASSUMPTIONS

The assumptions of the linear regression models for the admission duration outcome were assessed visually. For the fully adjusted models (Model 4) for the three age groups, the distributions of the residuals were checked by plotting histograms of the studentised residuals. As can be seen in Figures A.6.13 to A.6.15, the residuals were approximately normally distributed, and thus the linear regression assumption of normality of residuals was approximately met.

Additionally, equality of variance (homoscedasticity) was checked by plotting the residuals against the fitted values to see how the residuals varied as the fitted values increased (Figures A.6.16 to A.6.18). The plots indicate that the variance of the residuals was approximately constant as the fitted values increased, suggesting that the assumption of homoscedasticity was met.

Figures A.6.13 to A.6.15 Histograms of residuals from linear regression models for log admission days per admission outcome (fully adjusted Model 4)

Figure A.6.13 Children aged 0–14 years

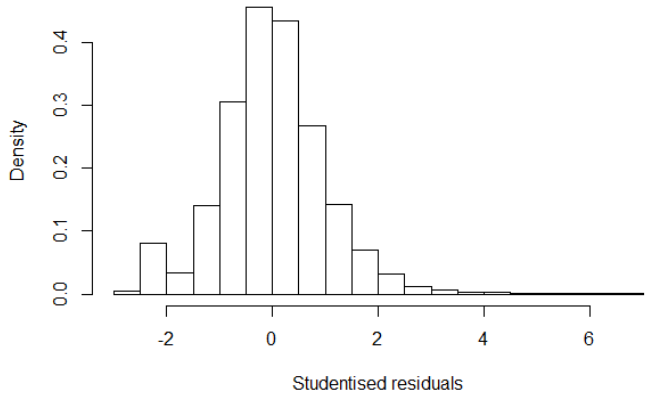


Figure A.6.14 Adults aged 15–64 years

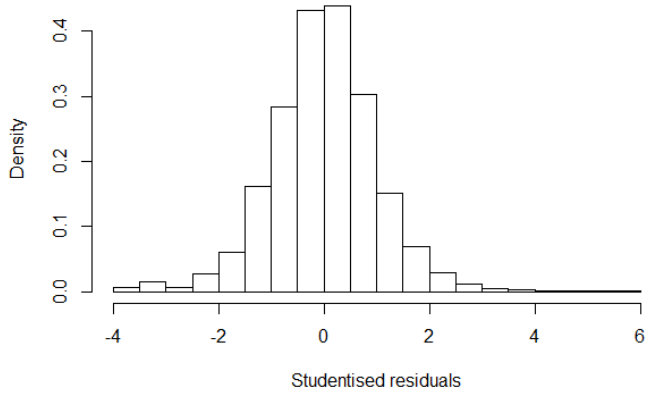
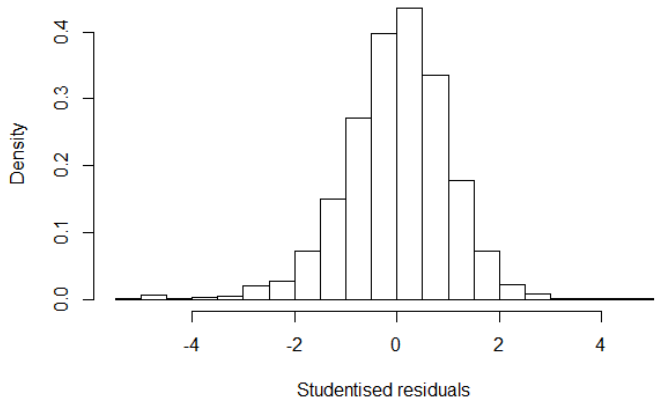


Figure A.6.15 Adults aged 65+ years



Figures A.6.16 to A.6.18 Plots of residuals versus fitted values from linear regression models for log admission days per admission outcome (fully adjusted Model 4)

Figure A.6.16 Children aged 0–14 years

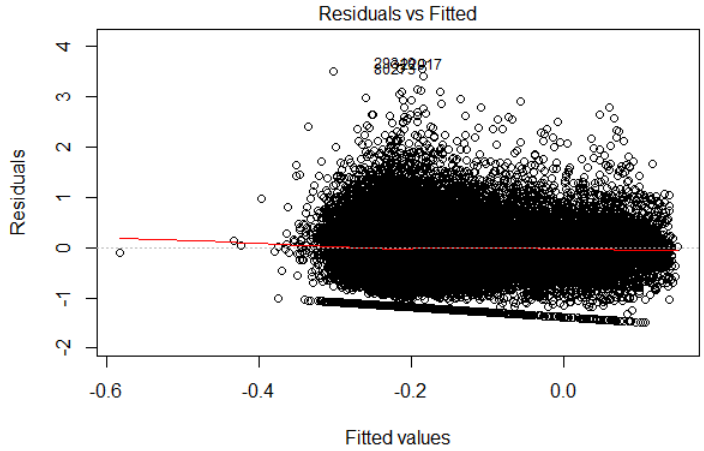


Figure A.6.17 Adults aged 15–64 years

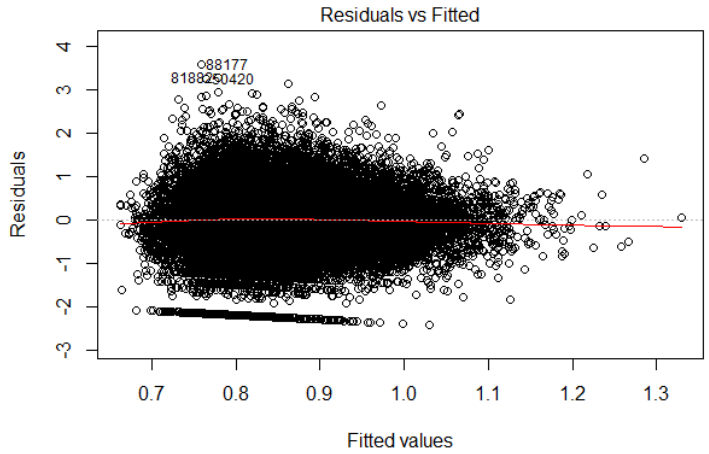
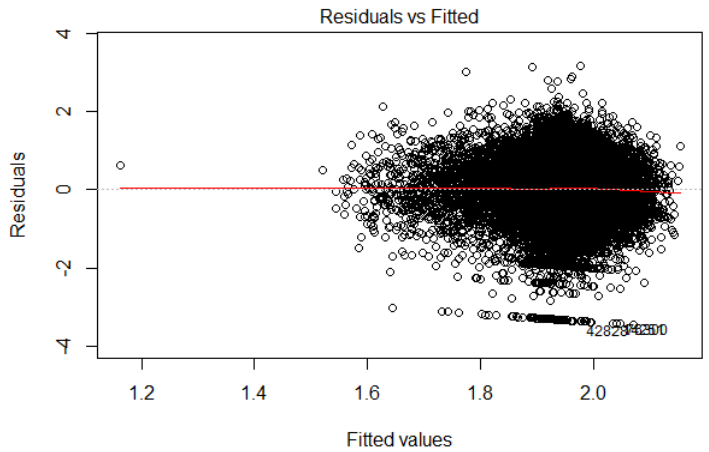


Figure A.6.18 Adults aged 65+ years



A.6.3 Analysis with more specific definition of IID

To investigate the potential impact of including the ICD-10 code K52.9 (unspecified non-infective gastroenteritis and colitis) in the definition of an emergency hospital admission for IID, the main analysis was repeated using a different definition of IID, that excluded codes K52.9 and A09.9 (gastroenteritis and colitis of unspecified origin). This definition of IID was more specific, but was probably less sensitive, given that possible cases of IID were likely excluded.

The multivariate regression results using this more specific definition are displayed in Tables A.6.1 to A.6.4. The results were similar to those observed in the main analysis. Increasing neighbourhood income deprivation was statistically significantly associated with increasing emergency hospital admission rates for IID and increasing duration of admissions, for all ages.

Table A.6.1 Multivariate negative binomial regression models for IID emergency hospital admission rates for English LSOAs, for children aged 0–14 years

	IID emergency hospital admission rates for children aged 0–14 years							
	IRR (95% CI)							
	Model 1		**Model 2**		**Model 3**		**Model 4**	
Proportion White (≤70%)	ref		ref		ref		ref	
>70% to ≤90%	1.06	(1.03-1.09)*	1.26	(1.23-1.29)***	1.26	(1.23-1.29)***	1.28	(1.25-1.32)***
>90%	1.23	(1.20-1.25)***	1.53	(1.50-1.57)***	1.52	(1.49-1.56)***	1.68	(1.64-1.72)***
Income deprivation (Q1)			ref		ref		ref	
Q2			1.21	(1.18-1.25)***	1.20	(1.17-1.23)***	1.21	(1.18-1.24)***
Q3			1.46	(1.43-1.50)***	1.44	(1.40-1.48)***	1.45	(1.41-1.49)***
Q4			1.77	(1.72-1.81)***	1.72	(1.67-1.77)***	1.70	(1.65-1.75)***
Q5 (most deprived)			2.20	(2.14-2.25)***	2.13	(2.07-2.19)***	2.06	(2.00-2.12)***
Proportion long-term health problem (Q1)					ref		ref	
Q2					1.07	(1.05-1.10)**	1.08	(1.05-1.10)**
Q3					1.08	(1.05-1.11)**	1.08	(1.06-1.11)**
Q4 (greatest proportion)					1.07	(1.04-1.10)*	1.07	(1.04-1.10)**
Distance GP (km)							1.01	(1.00-1.02)*
Distance hospital (km)							0.99	(0.99-0.99)***
Classification (Urban)							ref	
Rural							0.95	(0.93-0.98)*
Log-likelihood	-87395.6		-85512.1		-85489.2		-85182	
AIC	174799.3		171040.1		171000.5		170392	
BIC	174832.8		171107.1		171092.6		170509.3	
Number LSOAs	32011		32011		32011		32011	

CI = confidence interval; GP = general practice; IID = infectious intestinal disease; IRR = incident rate ratio; km = kilometre; LSOA = Lower Super Output Area; ref = reference category
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

Table A.6.2 Multivariate negative binomial regression models for IID emergency hospital admission rates for English LSOAs, for adults aged 15–64 years

	IID emergency hospital admission rates for adults aged 15–64 years							
	IRR (95% CI)							
	Model 1		**Model 2**		**Model 3**		**Model 4**	
Proportion White (≤70%)	ref		ref		ref		ref	
>70% to ≤90%	0.87	(0.85-0.90)***	1.02	(0.99-1.04)	0.98	(0.96-1.01)	0.99	(0.97-1.02)
>90%	0.88	(0.86-0.90)***	1.08	(1.06-1.11)**	0.97	(0.95-1.00)*	1.03	(1.00-1.06)*
Income deprivation (Q1)			ref		ref		ref	
Q2			1.12	(1.09-1.16)***	1.04	(1.01-1.08)*	1.05	(1.02-1.08)*
Q3			1.25	(1.21-1.28)***	1.07	(1.04-1.11)*	1.08	(1.05-1.12)**
Q4			1.49	(1.45-1.53)***	1.16	(1.12-1.20)***	1.16	(1.12-1.20)***
Q5 (most deprived)			1.87	(1.82-1.92)***	1.30	(1.25-1.35)***	1.29	(1.23-1.34)***
Proportion long-term health problem (Q1)					ref		ref	
Q2					1.17	(1.14-1.20)***	1.17	(1.14-1.20)***
Q3					1.30	(1.26-1.33)***	1.30	(1.26-1.34)***
Q4 (greatest proportion)					1.52	(1.47-1.58)***	1.52	(1.46-1.57)***
Distance GP (km)							1.01	(1.01-1.02)*
Distance hospital (km)							0.99	(0.99-0.99)***
Classification (Urban)							ref	
Rural							1.01	(0.98-1.03)
Log-likelihood	-62748.9		-61474.9		-61225.2		-61083.8	
AIC	125505.8		122965.8		122472.4		122195.6	
BIC	125539.4		123033		122564.8		122313.2	
Number LSOAs	32836		32836		32836		32836	

CI = confidence interval; GP = general practice; IID = infectious intestinal disease; IRR = incident rate ratio; km = kilometre; LSOA = Lower Super Output Area; ref = reference category
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

Table A.6.3 Multivariate negative binomial regression models for IID emergency hospital admission rates for English LSOAs, for adults aged 65+ years

	IID emergency hospital admission rates for adults aged 65+ years							
	IRR (95% CI)							
	Model 1		**Model 2**		**Model 3**		**Model 4**	
Proportion White (≤70%)	ref		ref		ref		ref	
>70% to ≤90%	0.99	(0.94-1.04)	1.11	(1.05-1.16)*	1.13	(1.07-1.19)*	1.13	(1.07-1.19)*
>90%	0.80	(0.76-0.83)***	1.01	(0.97-1.06)	1.02	(0.98-1.07)	1.10	(1.05-1.15)*
Income deprivation (Q1)			ref		ref		ref	
Q2			1.18	(1.15-1.22)***	1.14	(1.11-1.18)***	1.17	(1.13-1.20)***
Q3			1.37	(1.33-1.41)***	1.27	(1.23-1.31)***	1.30	(1.26-1.34)***
Q4			1.63	(1.58-1.67)***	1.44	(1.39-1.49)***	1.44	(1.39-1.48)***
Q5 (most deprived)			2.10	(2.04-2.16)***	1.75	(1.69-1.82)***	1.72	(1.66-1.78)***
Proportion long-term health problem (Q1)					ref		ref	
Q2					1.07	(1.05-1.10)**	1.05	(1.03-1.08)*
Q3					1.14	(1.11-1.18)***	1.09	(1.06-1.12)**
Q4 (greatest proportion)					1.28	(1.24-1.33)***	1.21	(1.17-1.25)***
Distance GP (km)							0.98	(0.98-0.99)*
Distance hospital (km)							0.99	(0.99-0.99)***
Classification (Urban)							ref	
Rural							0.92	(0.90-0.95)**
Log-likelihood	-60680.5		-59149.5		-59036.5		-58781.3	
AIC	121369		118315.1		118095		117590.6	
BIC	121402.2		118381.3		118186		117706.6	
Number LSOAs	29163		29163		29163		29163	

CI = confidence interval; GP = general practice; IID = infectious intestinal disease; IRR = incident rate ratio; km = kilometre; LSOA = Lower Super Output Area; ref = reference category
*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

Table A.6.4 Multivariate linear regression models for log IID admission days per admission for English LSOAs, stratified by age

	Log IID admission days per admission‡											
	Children aged 0–14 years				Adults aged 15–64 years				Adults aged 65+ years			
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 1**	**Model 2**	**Model 3**	**Model 4**
Proportion White (%)†	-0.0433***	-0.0404***	-0.0402***	-0.0377***	-0.0197**	-0.0101*	-0.0201**	-0.0187**	0.0340**	0.0470***	0.0465***	0.0519***
std. error	0.0016	0.0017	0.0017	0.0018	0.0030	0.0032	0.0036	0.0037	0.0059	0.0060	0.0060	0.0062
Income deprivation (%)†		0.0188**	0.0206*	0.0179*		0.0540***	-0.0021	-0.0013		0.0635***	0.0808***	0.0764***
std. error		0.0037	0.0043	0.0043		0.0057	0.0111	0.0112		0.0067	0.0095	0.0095
Proportion long term health problem (%)†			-0.0238	-0.0239			0.1368**	0.1366**			-0.0234*	-0.0290*
std. error			0.0263	0.0263			0.0231	0.0232			0.0090	0.0091
Distance GP (km)				-0.0052				0.0064				-0.0240*
std. error				0.0031				0.0048				0.0052
Distance hospital (km)				-0.0012*				-0.0022*				-0.0007
std. error				0.0005				0.0008				0.0009
Rural [a]				-0.0014				0.0180				0.0153
std. error				0.0126				0.0200				0.0218
Log-likelihood	-29262.9	-29250.2	-29249.8	-29242.3	-38422.3	-38378	-38360.5	-38356.1	-36642.3	-36598	-36594.6	-36580.2
AIC	58531.8	58508.5	58509.7	58500.6	76850.6	76764	76731	76728.2	73290.7	73204	73199.3	73176.5
BIC	58556.7	58541.8	58551.3	58567.1	76875.3	76797	76772.2	76794.1	73315.1	73236.6	73239.9	73241.5
Number LSOAs	30215	30215	30215	30215	27969	27969	27969	27969	25059	25059	25059	25059

GP = general practice; IID = infectious intestinal disease; km = kilometre; LSOA = Lower Super Output Area; std. error = standard error

[a] Reference category = Urban

*p <0.05 , **p <0.1^{-5} , ***p <0.1^{-10}

‡ Regression coefficients and standard errors displayed in table

† Variables entered into model in units of 10% point

Appendices to Chapter 7

The appendix for Chapter 7 features the publications and presentations that have arisen from this thesis thus far.

Articles

Rose TC*, Adams NL*, Taylor-Robinson DC, Barr B, Hawker J, O'Brien SJ, Violato M, Whitehead M (2016) Relationship between socioeconomic status and gastrointestinal infections in developed countries: a systematic review protocol. *Systematic Reviews*, 5:13.

Rose TC, Adams NL, Barr B, Hawker J, O'Brien SJ, Violato M, Whitehead M, Taylor-Robinson DC (2017) Socioeconomic status is associated with symptom severity and sickness absence in people with infectious intestinal disease in the UK. *BMC Infectious Diseases*, 17:447.

Abstracts and presentations

Rose TC*, Adams NL*, Taylor-Robinson DC, Barr B, Hawker J, O'Brien SJ, Whitehead M (2015) *Are there socioeconomic inequalities in the risks and consequences of gastrointestinal infections?* Presented at the NIHR HPRU in Gastrointestinal Infections Annual Scientific Conference, 25th March: London, UK.

Rose TC, Adams NL, Taylor-Robinson DC, Barr B, Hawker J, O'Brien SJ, Violato M, Whitehead M (2016) Relationship between socioeconomic status and measures of infectious intestinal disease severity. *European Journal of Public Health*, 26(suppl 1):ckw166.060.

Rose TC, Adams NL, Taylor-Robinson DC, Barr B, Hawker J, O'Brien SJ, Violato M, Whitehead M (2016) *Does socioeconomic status influence disease severity and sickness absence in people with infectious intestinal disease?* Presented at the NIHR HPRU in Gastrointestinal Infections Annual Scientific Conference, 14th March: Oxford, UK.

Rose TC*, Adams NL*, Taylor-Robinson DC, Barr B, Hawker J, O'Brien SJ, Violato M, Whitehead M (2017) *Relationship between socioeconomic status and risk of gastrointestinal infections in developed countries: a systematic review and meta-analysis*. Presented at the NIHR HPRU in Gastrointestinal Infections Annual Scientific Conference, 1st March: Liverpool, UK.

*Joint first author

Rose et al. Systematic Reviews (2016) 5:13
DOI 10.1186/s13643-016-0187-7

Systematic Reviews

PROTOCOL Open Access

Relationship between socioeconomic status and gastrointestinal infections in developed countries: a systematic review protocol

Tanith C. Rose[1,3,7*], Natalie Adams[1,3,5], David C. Taylor-Robinson[1,3], Benjamin Barr[1,3], Jeremy Hawker[1,5], Sarah O'Brien[1,4], Mara Violato[2,6] and Margaret Whitehead[1,3]

Abstract

Background: The association between low socioeconomic status (SES) and poor health is well documented in the existing literature. Nonetheless, evidence on the relationship between SES and gastrointestinal (GI) infections is limited, and the mechanisms underlying this relationship are not well understood with published studies pointing to conflicting results. This review aims to identify studies that investigate the relationship between SES and GI infections in developed countries, in order to assess the direction of the association and explore possible explanations for any differences in the risk, incidence or prevalence of GI infections across socioeconomic groups.

Methods: Three systematic methods will be used to identify relevant literature: electronic database, reference list and grey literature searching. The databases MEDLINE, Scopus and Web of Science Core Collection will be searched using a broad range of search terms. Screening of the results will be performed by two reviewers using pre-defined inclusion and exclusion criteria. The reference lists of included studies will be searched, and Google will be used to identify grey literature. Observational studies reporting quantitative results on the prevalence or incidence of any symptomatic GI infections by SES, in a representative population sample from a member country of the Organisation for Economic Co-operation and Development (OECD), will be included. Data will be extracted using a standardised form. Study quality will be assessed using the Liverpool University Quality Assessment Tools (LQAT). A narrative synthesis will be performed including tabulation of studies for comparison.

Discussion: This systematic review will consolidate the existing knowledge on the relationship between SES and GI infections. The results will help to identify gaps in the literature and will therefore provide an evidence base for future empirical studies to deepen the understanding of the relationship, including effective study design and appropriate data analysis methods. Ultimately, gaining insight into this relationship will help to inform policies to reduce any health inequalities identified.

Systematic review registration: PROSPERO CRD42015027231

Keywords: Socioeconomic factors, Income, Social class, Employment, Education, Gastrointestinal infection, Diarrhoea, Gastroenteritis, Foodborne diseases

* Correspondence: Tanith.Rose@liverpool.ac.uk
Tanith C. Rose and Natalie Adams are joint first authors.
[1]NIHR Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, Liverpool L69 7BE, UK
[3]Department of Public Health and Policy, University of Liverpool, Liverpool, UK
Full list of author information is available at the end of the article

Background

There is strong evidence of a social gradient in most health outcomes whereby the poorest in society experience greater levels of illness and premature death than those further up the socioeconomic scale [1]. Socioeconomic inequalities are linked to both causes and consequences of ill health [2] and have been well documented in diseases of a non-infectious nature, such as coronary heart disease and cancer [3]. Whilst there is evidence that the incidence of many infectious diseases, such as tuberculosis and human immunodeficiency virus [4–6], varies by social group, the association between socioeconomic status (SES) and gastrointestinal (GI) infections in particular is not well understood.

Gastrointestinal infections, caused by organisms such as bacteria, viruses or protozoa, are a common source of disease in the UK, leading to diarrhoea and vomiting and potentially more serious health problems, all of which can interfere with normal daily life. Previous studies have estimated that around 25 % of people in the UK will suffer an episode of infectious intestinal disease (IID) per year and that foodborne illness (a proportion of IID) in England and Wales costs around £1.5 billion per annum [7, 8]. It is reported that 10 % of children present to healthcare services with gastroenteritis each year, accounting for 16 % of paediatric accident and emergency presentations in one study [9]. There are eight million absences from school and at least 11 million working days lost to the economy each year due to GI infections [7].

The impact of SES on vulnerability to GI infections is unclear, and the limited existing evidence points to conflicting results. Higher prevalence of GI infections is often thought to be associated with more advantaged individuals. However, a recent systematic review looking at the impact of SES on laboratory-confirmed foodborne illness in developed countries suggests that this relationship is not so clear [10]. Newman et al. [10] identified 16 studies across four pathogens with mixed results, differing by pathogen. For example, in the most disadvantaged populations compared to the least disadvantaged, *Listeria* was more common, but *Campylobacter* was less common. In addition to the papers identified by Newman et al. [10], inconsistent results have also been observed among studies that have used syndromic definitions of GI infections, with some reporting higher rates of GI infections among those in lower socioeconomic groups [4, 11, 12] and others observing the opposite [13, 14]. These results clearly demonstrate the disagreements within this area of research.

A number of factors could explain these inconsistent results. The studies identified thus far cover a broad range of pathogens, and it may be that the relationship differs depending on whether the data are analysed at an all-GI-infection, pathogen-specific or species-specific level. This might suggest that the mode of transmission of an organism plays a role in the relationship and that this could be related to potentially socially patterned risk such as rural versus urban residency or exotic foreign travel. Furthermore, these studies have used different study designs, measured SES and GI infection in different ways and controlled for various confounders (such as age, labour market attachment, country of birth and agricultural occupation). Appropriate adjustment of confounding variables requires an understanding of the underlying mechanisms linking SES to GI infection risk, but there is little empirical evidence in this area.

A systematic review is warranted to summarise, organise and make sense of the contradictory findings observed in the literature. Our review aims to build on previous work by exploring the relationship between SES and a full range of GI infections. As it is possible that various socioeconomic or healthcare-seeking behavioural factors could influence whether an individual is diagnosed with a GI infection, we have also included syndromic definitions of GI infections. We aim to explore the current knowledge of the relationship in developed countries; assess the magnitude, statistical significance and direction of the association; and shed light into possible explanations for any observed differences in the risk, incidence or prevalence of GI infections across socioeconomic groups. The results of this review will help to inform the development of empirical research projects by identifying gaps in the literature and areas where further research is required. It will provide evidence of the methods employed previously to investigate the relationship between SES and GI infections, including information on the relevant confounding variables used.

Methods/design

To improve the transparency and completeness of the protocol, a completed copy of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols 2015 (PRISMA-P 2015) checklist [15] can be found in Additional file 1.

Research question

For individuals from developed countries, is lower compared to higher SES associated with the incidence or prevalence of GI infection?

Population

Any individual, of any age or gender, from a developed country will be included. A developed country is defined as being a member country of the Organisation for Economic Co-operation and Development (OECD). The OECD aims to continually monitor the economic

developments of its 34 member countries and provides policy recommendations to help governments tackle poverty through economic growth and stability [16].

Exposure

The exposure of interest is lower compared to higher SES, measured at the individual or aggregate level by income, education, occupation, employment or deprivation of area of residence.

Outcome

The primary outcome of interest will be the incidence or prevalence of any symptomatic GI infection measured using population level surveys, routine surveillance systems, laboratory data or hospitalisation data and includes syndromic definitions of GI infections without a laboratory diagnosis.

Inclusion/exclusion criteria

Observational studies (cross-sectional, ecological, case-control, cohort [prospective and retrospective]) reporting quantitative results and analysis of empirical data on the prevalence or incidence of any symptomatic GI infection by SES, in a representative population sample, will be included. Socioeconomic status can be measured by occupation, income, education, employment or deprivation at the individual or aggregate level. Only studies conducted in developed countries (defined as being a member country of the OECD), written in or translated into English, reporting on human subjects and using data collected after 1980, will be included. For countries that joined the OECD after 1980, data collection must have occurred after the date the country became a member of the OECD. Studies not meeting the above criteria, including case studies, case series or literature reviews, or studies reporting on outbreaks of GI infection, travel-associated illness only or asymptomatic infections only will be excluded. Studies conducted solely in a specific population subgroup without a general population comparator group or studies conducted in institutional settings such as nurseries, hospitals or the military will be excluded.

Search strategy

Three search strategies will be used to identify as much relevant literature as possible. Firstly, the electronic searching of three databases will be performed: MEDLINE (Ovid), Scopus and Web of Science Core Collection. The choice of database was discussed with a university librarian, and the three databases chosen were considered most relevant to the research question and likely to yield the highest number of relevant papers.

The search terms were piloted prior to selection and are comprised of specific GI infection and symptom-based terms, socioeconomic and inequality terms, and developed countries of interest (Additional file 2). Relevant synonyms for the SES and GI infection terms were identified using Roget's Thesaurus online [17] and the thesaurus in MEDLINE by mapping and inspecting the tree for each term. Relevant terms mentioned in articles identified in a pilot search of the literature were also added. Ultimately, the GI infection terms were selected because they represent the main GI pathogens known to cause the greatest burden to public health in the developed world. Whilst not exhaustive, the list is intended to provide a broad spectrum of bacterial, viral and protozoal infections.

The search terms for MEDLINE were developed initially. Where possible, terms were exploded to broaden the search. Terms were added as keywords if they could not be exploded or if the exploded terms were not relevant to the research question. Truncation and proximity operators were also applied as necessary to broaden the search. Terms were combined using Boolean operators.

For consistency, the exact same terms were used for Scopus and Web of Science Core Collection; however, as the functionality of each database is different, it was necessary to adapt the terms developed in MEDLINE for correct use in Scopus and Web of Science Core Collection. Specifically, the terms contained within the exploded terms in MEDLINE needed to be added as individual search terms for use in Scopus and Web of Science Core Collection, and it was necessary to indicate phrases with quotation marks. Additionally, the proximity operators differed for each database.

When the searches are run in Scopus and Web of Science Core Collection, each term will be searched for within the title, abstract and keywords of the documents contained in each database. Filters within the three databases will be applied to restrict the results to publications that have used data from 1980 to the present. As social conditions within countries change over time through development, and methods of classifying SES are also modified over time, restricting to publications using data from 1980 onwards will ensure that the results are as relevant as possible to the present day. Results will also be limited to publications available in the English language. Additionally, where available, filters for 'human subjects' and 'document type' will be applied to the database search results. All of these filters directly relate to the inclusion criteria. The publications remaining after the filters are applied will then be exported into reference managing software. In this software, the publications from the three databases will be combined and duplicates removed. The remaining publications will then be screened for relevance using the inclusion and exclusion criteria.

Titles and abstracts of the publications will be screened independently by two authors (NA and TR) to

ensure consistency in the application of the inclusion and exclusion criteria. Any discrepancies will be discussed and re-examined until an agreement is reached between both reviewers. The full text for studies deemed relevant after title and abstract screening will be retrieved and reviewed in the same way. Where full texts are not available, they will be sought via institutional library sharing agreements. All full-text studies will be screened independently by the same two reviewers to ensure that they conform to the inclusion and exclusion criteria.

The second strategy will consist of searching the reference lists of any studies selected for inclusion in the final review to identify potentially relevant articles that may have been missed by the electronic database searches. The abstracts of any references considered potentially relevant will be sought and screened for inclusion using the pre-defined inclusion and exclusion criteria. The full text for studies deemed relevant after title and abstract screening will be retrieved and reviewed in the same way. This reference list search will be conducted independently by two reviewers (NA and TR), and discrepancies will be discussed and eventually agreed upon at each stage.

The third method will be to conduct a search of the grey literature by entering the terms 'gastrointestinal infection', 'gastroenteritis', 'diarrhoea', 'diarrhea', 'socioeconomic', 'social class', 'income' and 'deprivation' into the Google internet search engine and the Google Scholar search application and assessing the first 100 results. Each result will be inspected for relevance using the inclusion an exclusion criteria. Again, this will be performed independently by the two reviewers (NA and TR), and disagreement will be resolved through discussion.

Quality assessment

Risk of bias and quality assessment of the identified studies will be conducted by the review team, independently and then reconciled. The Liverpool University Quality Assessment Tool (LQAT) will be used for this review, which will allow the methodological quality of the studies to be assessed using a tool specific to each study design [18]. It incorporates a star rating system to assess and qualify absence of bias, misclassification and confounding. The LQAT has been used in previous systematic reviews [19, 20] and has been independently evaluated against other quality assessment tools [21]. Any discrepancies between reviewers in the quality assessment of the studies will be discussed and re-examined.

Data analysis and synthesis

To organise these data and to facilitate comparison, tables will be created by extracting data from each study into a standardised Excel spreadsheet. Data to be extracted will include the following: aim/hypothesis, study design, level of analysis, country, sample size, age, age category, type of GI infection, GI infection method of measurement and data source, measure of SES, SES method of measurement and data source, covariates, statistically significant results, non-significant results, conclusions and quality assessment. Extracted data will be checked for accuracy by at least one other reviewer.

Due to the broad scope of this review, it is anticipated that there will be considerable heterogeneity between studies in terms of design, populations studied and the measurement of primary exposures and outcomes. The synthesis strategy will be driven by the data available; however, to explore the relationship between GI infections and SES, it is anticipated that a subgroup analysis will be performed on study design factors and potential moderating factors of the relationship, including but not limited to the following: pathogen type (based on mode of transmission); age; country (based on climate and relative level of development); methods used to measure GI infection; methods used to measure SES; and level of analysis (aggregate or individual). Separate tables will be created to compare and contrast the results of studies within and between the subgroups. If the data allow, further grouping of the studies within the subgroups will be performed to help summarise the study findings and answer the research question. The LQAT results will be used to determine the strength of the evidence from individual studies, and greater weight will be given to conclusions drawn from the most methodologically robust and reliable studies. A narrative synthesis will help to make sense of what is anticipated to be a diverse body of evidence and may lead to potential explanations for the contrasting findings observed in the literature. The methods used will be written up transparently, and the robustness of the synthesis will be assessed [22].

Where homogenous data allow, meta-analyses will be conducted on combined results. The synthesis strategy outlined above will assist in identifying data suitable for meta-analysis. Heterogeneity will be assessed by examining the forest plots to detect overlapping confidence intervals, using the chi^2 test with a P value of 0.10 to indicate statistical significance, and also applying the I^2 statistic with values of 30 to 60 %, 50 to 90 % and 75 to 100 % used to denote moderate, substantial and considerable levels of heterogeneity, respectively [23]. If the data allow, publication bias will be assessed using a funnel plot, and sensitivity analysis on the basis of study quality will be conducted to explore the robustness of the meta-analysis. RevMan software will be used to conduct these analyses [24]. A 'Summary of findings' table [25] will be used to present the results, and the Grading of Recommendations, Assessments, Development and Evaluation approach will be used to assess the quality of the body of evidence [26].

Dissemination

The systematic review will be submitted for publication. The findings of the review and data will be presented at conferences and will contribute to two PhD projects as part of the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections [27].

Discussion

Our systematic review aims to provide new insight into the understanding of the mixed results on the relationship between SES and GI infections as suggested by Newman et al. [10], by broadening the focus to a wider range of symptomatic GI infections and exploring whether a more conclusive pattern can be identified. This includes syndromic definitions of GI infections in the absence of laboratory confirmation. By including these definitions, we aim to identify literature on the burden of symptoms by SES and attempt to capture population groups who may not seek healthcare for their illness and consequently may not be included in studies which use laboratory data to identify cases only. This is particularly important for this review as the decision to seek healthcare may be related to SES.

In the UK, it is estimated that 17 million cases of infectious intestinal disease occur every year, resulting in approximately one million general practice consultations [7]. This, coupled with an increasingly overburdened National Health Service (NHS), highlights the importance of understanding the role of SES in GI infections in order to devise policies to target the strata of the population most at risk.

The results of this review will provide a more comprehensive evidence base of the relationship between symptomatic GI infections and SES to inform the development of empirical studies, including effective study design and appropriate data analysis methods, which will be used in two PhD projects.

Additional files

Additional file 1: PRISMA-P 2015 checklist. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols 2015 (PRISMA-P 2015) checklist was used to develop this protocol. Items 1b and 4 were not applicable.

Additional file 2: Search terms for MEDLINE, Scopus and Web of Science Core Collection. The search terms that will be used to identify relevant literature across three databases.

Abbreviations

GI: gastrointestinal; IID: infectious intestinal disease; LQAT: Liverpool University Quality Assessment Tools; NHS: National Health Service; NIHR HPRU: National Institute for Health Research Health Protection Research Unit; OECD: Organisation for Economic Co-operation and Development; PRISMA-P 2015: Preferred Reporting Items for Systematic Reviews and Meta-analyses for Protocols 2015; SES: socioeconomic status.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NA and TR wrote the protocol. DTR, BB, JH, SOB, MV and MW conceived the initial idea for the study, critically appraised the protocol and also contributed to its development by revising different versions. All authors approved the final version and take responsibility for its content.

Authors' information

NA and TR are PhD students funded by the National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections at the University of Liverpool.

Funding

This research is funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at the University of Liverpool in partnership with Public Health England (PHE), University of East Anglia, University of Oxford and the Institute of Food Research. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

Author details

[1]NIHR Health Protection Research Unit in Gastrointestinal Infections, University of Liverpool, Liverpool L69 7BE, UK. [2]NIHR Health Protections Research Unit in Gastrointestinal Infections, University of Oxford, Oxford, UK. [3]Department of Public Health and Policy, University of Liverpool, Liverpool, UK. [4]Institute of Infection and Global Health, University of Liverpool, Liverpool, UK. [5]Public Health England, London, UK. [6]Health Economics Research Centre, University of Oxford, Oxford, UK. [7]Department of Public Health and Policy, Institute of Psychology, Health and Society, University of Liverpool, Whelan Building, Liverpool L69 3GB, UK.

Received: 21 December 2015 Accepted: 12 January 2016
Published online: 21 January 2016

References

1. Wilkinson R, Marmot M. Social determinants of health: the solid facts. 2nd ed. Geneva: World Health Organisation; 2003. http://www.euro.who.int/__data/assets/pdf_file/0005/98438/e81384.pdf. Accessed 17 Feb 2015.
2. Whitehead M, Dahlgren G. Concepts and principles for tackling social inequities in health: levelling up part 1. Copenhagen: World Health Organisation Regional Office for Europe; 2006. http://www.enothe.eu/cop/docs/concepts_and_principles.pdf. Accessed 17 Feb 2015.
3. Graham H. Understanding health inequalities. Maidenhead: Open University Press, McGraw-Hill Education; 2009.
4. Biering-Sorensen S, Sondergaard G, Vitting Andersen K, Andersen AM, Mortensen LH. Time trends in socio-economic factors and risk of hospitalisation with infectious diseases in pre-school children 1985–2004: a Danish register-based study. Paediatr Perinat Epidemiol. 2012;26(3):226–35.
5. Hughes GJ, Gorton R. Inequalities in the incidence of infectious disease in the North East of England: a population-based study. Epidemiol Infect. 2015;143(1):189–201.
6. Semenza JC. Strategies to intervene on social determinants of infectious diseases. Euro Surveill. 2010;15(27):32–9.
7. Food Standards Agency. The second study of infectious intestinal disease in the community (IID2 Study). https://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/b18021. Accessed 15 Dec 2015.
8. Food Standards Agency. Foodborne disease strategy 2010–15: an FSA programme for the reduction of foodborne disease in the UK. 2011. http://www.food.gov.uk/sites/default/files/multimedia/pdfs/fds2015.pdf. Accessed 16 Dec 2015.
9. Armon K, Stephenson T, Gabriel V, MacFaul R, Eccleston P, Werneke U, et al. Determining the common medical presenting problems to an accident and emergency department. Arch Dis Child. 2001;84(5):390–2.
10. Newman KL, Leon JS, Rebolledo PA, Scallan E. The impact of socioeconomic status on foodborne illness in high-income countries: a systematic review. Epidemiol Infect. 2015;143(12):2473–85.
11. Olowokure B, Hawker J, Weinberg J, Gill N, Sufi F. Deprivation and hospital admission for infectious intestinal diseases. Lancet. 1999;353(9155):807–8.

12. Pockett RD, Adlard N, Carroll S, Rajoriya F. Paediatric hospital admissions for rotavirus gastroenteritis and infectious gastroenteritis of all causes in England: an analysis of correlation with deprivation. Curr Med Res Opin. 2011;27(4):777–84.
13. de Wit MAS, Koopmans MPG, Kortbeek LM, Wannet WJB, Vinje J, Van Leusden F, et al. Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. Am J Epidemiol. 2001;154(7):666–74.
14. Pollard CM, Meng X, Williamson S, Dodds J, Binns CW. Eating out is associated with self-reported food poisoning: a Western Australia population perspective, 1998 to 2009. Public Health Nutr. 2014;17(10):2270–7.
15. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: elaboration and explanation. Br Med J. 2015;2015:1–25.
16. Organisation for Economic Co-operation and Development. What we do and how. http://www.oecd.org/about/whatwedoandhow/. Accessed 26 Oct 2015.
17. Thesaurus.com. http://www.thesaurus.com/. Accessed 15 Dec 2015.
18. Pope D. Introduction to systematic reviews [lecture]. Liverpool: University of Liverpool; 2015.
19. Puzzolo E, Stanistreet S, Pope D, Bruce NG, Rehfuess EA. Factors influencing the large scale uptake by households of cleaner and more efficient household energy technologies. A systematic review. Evidence for Policy and Practice Information and Co-ordinating Centre, 2013. http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3426. Accessed 20 Oct 2015.
20. Rehfuess E, Puzzolo E, Stanistreet S, Pope D, Bruce NG. Enablers and barriers to large-scale uptake of improved solid fuel stoves: a systematic review. Environ Health Perspect. 2014;122(2):120–30.
21. Voss PH, Rehfuess EA. Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. J Epidemiol Community Health. 2013;67(1):98–104.
22. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M. Guidance on the conduct of narrative synthesis in systematic reviews: a product of the ESRC Methods Programme. Version 1. Swindon: Economic and Social Research Council; 2006.
23. Identifying and measuring heterogeneity. http://handbook.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm. Accessed 17 Dec 2015.
24. Review Manager (RevMan) [Computer program]. Version 5.3. The Nordic Cochrane Centre, The Cochrane Collaboration; 2014
25. General template for 'Summary of findings' tables. http://handbook.cochrane.org/chapter_11/11_5_3_general_template_for_summary_of_findings_tables.htm. Accessed 17 Dec 2015.
26. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328(7454):1490.
27. National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections. http://www.hprugi.nihr.ac.uk/. Accessed 15 Dec 2015.

Rose et al. BMC Infectious Diseases (2017) 17:447
DOI 10.1186/s12879-017-2551-1

BMC Infectious Diseases

RESEARCH ARTICLE Open Access

Socioeconomic status is associated with symptom severity and sickness absence in people with infectious intestinal disease in the UK

Tanith C. Rose[1,2,5*], Natalie L. Adams[1,2,3], Benjamin Barr[1,2], Jeremy Hawker[1,3], Sarah J. O'Brien[1,2], Mara Violato[1,4], Margaret Whitehead[1,2] and David C. Taylor-Robinson[1,2]

Abstract

Background: The burden of infectious intestinal disease (IID) in the UK is substantial. Negative consequences including sickness absence are common, but little is known about the social patterning of these outcomes, or the extent to which they relate to disease severity.

Methods: We performed a cross-sectional analysis using IID cases identified from a large population-based survey, to explore the association between socioeconomic status (SES) and symptom severity and sickness absence; and to assess the role of symptom severity on the relationship between SES and absence. Regression modelling was used to investigate these associations, whilst controlling for potential confounders such as age, sex and ethnicity.

Results: Among 1164 cases, those of lower SES versus high had twice the odds of experiencing severe symptoms (OR 2.2, 95%CI;1.66–2.87). Lower SES was associated with higher odds of sickness absence (OR 1.8, 95%CI;1.26–2.69), however this association was attenuated after adjusting for symptom severity (OR 1.4, 95%CI;0.92–2.07).

Conclusions: In a large sample of IID cases, those of low SES versus high were more likely to report severe symptoms, and sickness absence; with greater severity largely explaining the higher absence. Public health interventions are needed to address the unequal consequences of IID identified.

Keywords: Socioeconomic factors, Occupation, Infectious intestinal disease, Diarrhoea, Sick leave, Symptom severity

Background

Infectious intestinal disease (IID) is extremely common, with an estimated 17 million sporadic cases occurring each year in the United Kingdom (UK) [1]. It also confers significant morbidity and associated healthcare costs. Around half of those who experience IID report absence from work or school which amounts to an estimated loss of nearly 19 million days per annum, with potential ramifications for adult earnings and child education [2]. Additionally, there are approximately one million general practice (GP) consultations for IID every year in the UK [1]. The burden of IID is clearly evident, yet relatively little is known about the extent of socioeconomic inequalities in the clinical, social and economic consequences of IID.

Studies conducted in developed countries, suggest individuals of low socioeconomic status (SES) compared to high, have higher rates of GP consultation [3, 4] and hospital admission due to IID [5–9]. For example, in the West Midlands in the UK, hospital admission rates for young children with IID were twice as high in the most deprived areas compared to the least [6]. However, the mechanisms explaining these apparent health inequalities are unknown. Contributing factors may include differential risk of infection, healthcare seeking behaviour, or disease severity across socioeconomic groups.

* Correspondence: Tanith.Rose@liverpool.ac.uk
Benjamin Barr and David C. Taylor-Robinson are senior authors.
[1]NIHR Health Protection Research Unit in Gastrointestinal Infections, Liverpool, UK
[2]Department of Public Health and Policy, University of Liverpool, Liverpool, UK
Full list of author information is available at the end of the article

Separating out the effects of these potential explanations is imperative to understand the role they play in generating the inequalities observed, and so that interventions and policies can be developed to tackle the problem. A cross-sectional analysis of IID cases identified in the English IID1 studies, showed that IID cases of lower SES (as measured by educational attainment) were more likely to present to their GP for an episode of IID, compared to those of higher SES [3]. In addition, disease severity was strongly predictive of GP presentation for IID, however numbers were insufficient to assess the relationship between SES and IID severity. These findings indicate that healthcare seeking behaviour for IID may be socially patterned, which potentially could be related to disease severity.

Negative consequences of IID also include sickness absence, which may or may not be related to IID severity. Rates of general (all cause) sickness absence, have been shown to be higher for those of lower SES compared to high [10], however some studies have demonstrated that this association can in part be explained by the increased levels of morbidity for those of lower SES [10, 11]. The few studies that have investigated the relationship between SES and sickness absence due to IID have produced conflicting results [12, 13]; and we are yet to find a study that has examined the role of IID severity on the relationship between SES and sickness absence. To gain a better understanding of inequalities in the consequences of IID, we analysed a large sample of IID cases to explore the association between SES and measures of self-reported IID symptom severity and sickness absence.

Methods

Study design and data source

We analysed cases of IID identified in the population-based IID2 study. The IID2 study was conducted across the UK in 2008–9 and contained several studies, the methods of which have been described in detail elsewhere [14]. The IID2 study was granted ethical approval by the North West Research Ethics Committee (07/MRE08/5) [15]. Participants gave written informed consent for their anonymised data to be used for future analyses.

The IID2 study contained two major components; a prospective cohort study and a GP presentation study. For the cohort study, patients were randomly selected from the registers of 88 general practices and invited to participate. Participants completed a baseline questionnaire containing questions on socio-demographic factors, and were followed-up weekly for one year to determine the incidence of IID. Incident cases completed symptom questionnaires including questions on symptom severity, absenteeism and recent foreign travel. For the GP presentation study, all patients who consulted their GP for an episode of IID across 37 of the 88 general practices, over a one year period were invited to participate in a survey which included the same socio-demographic and symptom questions as the former study.

Cases identified via both components of the IID2 study were combined for this analysis. Cases of IID were defined as people aged five years or older, with loose stools or clinically significant vomiting lasting less than two weeks, in the absence of a known non-infectious cause, preceded by a symptom-free period of three weeks [14]. We included cases aged five years or older, to limit potential misclassification of the more subjective symptoms, such as headache and nausea, in young children (see below details of symptom severity score). For cases meeting the case definition, all recurrent episodes of IID were removed regardless of the timeframe between episodes. If a case experienced more than one episode of IID during follow-up, only information related to the first episode was retained to create a sample of independent observations.

Outcomes and covariates

The outcomes of interest were symptom severity and sickness absence due to IID. The symptom severity score was derived from information on the presence/absence of nine symptoms, and the duration of four symptoms, which were self-reported by the cases, using previously published methods [3]. In brief, the presence and duration scores were multiplied, and the resulting product scores summed across the symptoms, creating an overall symptom severity score for each case (Additional file 1). The symptom severity variable was converted into tertiles, whereby three approximately equally sized groups were created according to the distribution of the severity score [3]. The second outcome of interest was sickness absence; a binary variable indicating whether the episode of IID prevented the case from going to work or school. Sickness absence was only defined for cases of school or working age (aged five years or older, and up to 60 years for women and 65 years for men, as older age groups were unlikely to be in work or education [16]).

The main exposure of interest was SES measured at the individual-level using the National Statistics Socioeconomic Classification (NS-SEC) [17]. The NS-SEC was designed to take into account the nature of modern inequalities, by measuring conditions of occupations and also employment relations [17, 18]. To derive the NS-SEC, participants answered via self-completion questionnaire, questions relating to the occupation and employment status of the main-earner in their household, reporting on the main-earner's current or last main job. Individuals were assigned the category 'Not classifiable' if information was missing and as such an NS-SEC class could not be calculated. We re-coded the five-class NS-SEC version to form the three-class version which can be assumed to have a

hierarchy [17]. The classes from high to low SES represented managerial/professional, intermediate and routine/manual occupations.

Potential confounding variables of the relationship between SES, symptom severity and sickness absence included in the analysis, were age, sex, ethnicity, foreign travel in the ten days before disease onset, and urban/rural residency (based on Super Output Areas) [19].

Statistical analysis

Ordinal logistic regression was employed for the symptom severity outcome, and logistic regression for the binary absence outcome. Model parameters were estimated by maximum likelihood. For the ordinal logistic regression models, the proportional odds assumption was assessed using graphical methods [20, 21]. Generalised additive models (GAMs) were used to assess the linear relationship between the continuous age variable and the outcomes (Additional file 1). There was a linear relationship between age and the log-odds of sickness absence, therefore age was included as a continuous variable when modelling the absence outcome. The relationship between age and symptom severity was non-linear, therefore a categorical age group variable was included when modelling symptom severity.

A hierarchical approach was used for the multivariate regression modelling. Firstly, we fitted baseline models for each of our two outcomes (symptom severity and sickness absence) with age, sex and ethnicity as independent variables. Secondly, we added NS-SEC as an additional independent variable to the models and tested the improvement in model fit using generalised likelihood ratio statistics to compare nested models. Thirdly, we tested whether the inclusion of additional confounders (recent foreign travel and urban/rural residency) improved the model fit. Finally, to explore whether differences in disease severity explained any association between NS-SEC and sickness absence, we added symptom severity as a control variable to the model with sickness absence as an outcome.

Listwise deletion was used as the method of handling missing data. For the two outcomes, cases with missing data within any of the variables to be included in the models were excluded. Sensitivity analyses were performed using multiple imputation by chained equations to impute missing data values for all of the variables included in the models.

We undertook several robustness tests, repeating our analyses using alternative cut-offs for the symptom severity categories; including recurrent episodes of IID within the same individual; using cases of all ages; and stratifying results by child and adult age groups. We also examined the appropriateness of combining cases from the IID2 cohort and GP presentation studies. Analyses were conducted using R (version 3.3.1).

Results

The IID2 studies identified 1915 cases meeting our inclusion criteria of which 1270 were of school or working age and included in the sickness absence analysis (see Additional file 1 for flow diagram). Characteristics of the cases stratified by NS-SEC are shown in Table 1. Around half of cases were in managerial/professional occupations, and the vast majority were of White ethnicity (>90%). Cases in routine/manual compared to managerial/professional occupations were less likely to reside in rural areas, be female or have travelled abroad before their illness. Age and ethnicity were not associated with NS-SEC.

Symptom severity

The symptom severity score ranged from 2 to 40 and was positively skewed. The boundaries for the tertiles were: mild (score 2–9), moderate (score 10–15) and severe (score 16–40). In total, 1164 (61%) cases had complete data for the variables of interest.

The univariate associations between symptom severity and the exposures are shown in Table 2, and two nested multivariate models for symptom severity are displayed in Table 3. The addition of NS-SEC to the baseline model improved the model fit when comparing the likelihoods of the models (Likelihood ratio χ^2 31.7; $P < 0.001$). For those in routine/manual compared to managerial/professional occupations the odds of experiencing severe IID symptoms, versus mild or moderate symptoms combined, were two times greater (OR 2.2, 95%CI;1.66–2.87). The odds of experiencing severe symptoms were greater for those of Non-White compared to White ethnicity, and for those aged 15–24 years compared to 5–14 years, however these estimates were based on small numbers (43 cases were of Non-White ethnicity; 61 cases were aged 15–24 years). There was no improvement in the model fit when the variables urban/rural residency and recent foreign travel were added to the Baseline + NS-SEC model, and therefore these models are not presented.

Sickness absence

Of the 1270 cases of school or working age, 818 (64%) had complete data for the variables of interest (Additional file 1). Over half of the cases (62%) were absent from work or school following their illness. Amongst the absentees, the majority took 1–2 days sick leave (62%), and few took more than five days (8%).

The univariate associations between sickness absence and the exposures are shown in Table 2, and three nested multivariate models for sickness absence are displayed in Table 4. The addition of NS-SEC produced a better fitting model compared to the baseline model (Likelihood ratio χ^2 10.2; $P = 0.006$). Those in routine/manual compared to managerial/professional occupations had a higher odds of absence (OR 1.8, 95%CI;1.26–2.69).

Table 1 Unadjusted distribution of each variable by NS-SEC category, for the two analysis samples (IID2 study 2008–9)

	Cases ≥5 years of age (*n* = 1915)					Cases school/working age (*n* = 1270)				
	Percentage within each category of NS-SEC					Percentage within each category of NS-SEC				
	Managerial/ professional (*n* = 949)	Intermediate (*n* = 330)	Routine/manual (*n* = 337)	*p*-value[a]	All cases[b] (*n* = 1915)	Managerial/professional (*n* = 662)	Intermediate (*n* = 215)	Routine/ manual (*n* = 228)	*p*-value[a]	All cases[b] (*n* = 1270)
Age group (years)										
5–14	12.2	9.4	9.5	0.342	10.6	17.5	14.4	14	0.337	16
15–24	4.4	5.2	7.4		5.2	6.3	7.9	11		7.9
25–44	24.1	22.7	23.7		22.9	34.6	34.9	35.1		34.5
45–64	36	36.7	33.5		35.1	41.5	42.8	39.9		41.7
65+	23.2	26.1	25.8		26.2					
Male	38	32.7	45.1	0.004	37.9	39.7	37.2	49.6	0.014	41
Ethnicity Non-White	3.4	4.8	3.6	0.468	4	3.9	7	4.8	0.185	5.1
Rural residence	30.6	30	19	<0.001	28.8	30.3	28.4	20.6	0.019	27.6
Travelled before illness	14.3	10.3	7.7	0.004	11.9	15.6	13.1	7.9	0.013	12.9
Symptom severity										
Mild	38.3	34.4	20.3	<0.001	24.5	40	36	20.7	<0.001	26.8
Moderate	34	33	35.5		24.9	36.7	32	38		27.1
Severe	27.7	32.6	44.2		23.4	23.3	32	41.3		22.1
Absent work/school	55.1	54.2	63	0.028	53.4	61.4	62.7	71.6	0.023	61

IID infectious intestinal disease, *NS-SEC* National Statistics Socioeconomic Classification, *SD* standard deviation
Figures expressed as percentages except where stated otherwise
[a]Statistical significance of relationship between NS-SEC and each variable, tested using χ^2 test
[b]Total number of cases includes those with missing NS-SEC
Missing data (%) for cases ≥5 years and cases school/working age, respectively: NS-SEC 16 and 13; Urban/rural 0.1 and 0.2; Foreign travel 0.6 and 0.3; Symptom severity 27 and 24; Absence 3.6 and 2.7

Table 2 Univariate associations for IID symptom severity and sickness absence outcomes (IID2 study 2008–9)

	Severe symptoms versus mild or moderate symptoms combined OR (95%CI)	Sickness absence versus no sickness absence OR (95%CI)[a]
	Cases with complete data ≥5 years of age (*n* = 1164)	Cases with complete data school/working age (*n* = 818)
Age group (years)		
5–14	reference	
15–24	2.88 (1.59–5.31)	
25–44	0.99 (0.68–1.45)	
45–64	0.70 (0.49–1.01)	
65+	0.60 (0.41–0.89)	
Age (years)		0.98 (0.97–0.99)
Sex		
Female	reference	reference
Male	0.92 (0.74–1.14)	0.94 (0.70–1.25)
Ethnicity		
White	reference	reference
Non-White	2.27 (1.28–4.10)	3.13 (1.38–8.41)
NS-SEC		
Managerial/ professional	reference	reference
Intermediate	1.21 (0.92–1.61)	1.13 (0.78–1.66)
Routine/manual	2.18 (1.67–2.86)	1.77 (1.22–2.58)
Residence		
Urban	reference	reference
Rural	0.82 (0.65–1.03)	0.98 (0.71–1.34)
Travelled before illness		
No	reference	reference
Yes	1.21 (0.88–1.66)	0.66 (0.44–0.99)
Symptom severity		
Mild		reference
Moderate		3.88 (2.75–5.51)
Severe		5.99 (4.07–8.95)

CI confidence interval, *IID* infectious intestinal disease, *NS-SEC* National Statistics Socioeconomic Classification, *OR* odds ratio
[a]Since the absence outcome was common, the odds ratios should not be interpreted as risk ratios

When symptom severity was added to this model the odds of absence for those in routine/manual compared to managerial/professional occupations was attenuated and rendered non-significant (OR 1.4, 95%CI;0.92–2.07). There was a dose-response relationship between symptom severity and the odds of absence. Those with severe compared to mild symptoms had five times the odds of absence (OR 5.3, 95%CI;3.54–7.93). Again, there was no improvement in the model fit when the variables urban/rural residency and recent foreign travel were added to the Baseline + NS-SEC model, and therefore these models are not presented.

Table 3 Multivariate models for severe IID symptoms, versus mild or moderate symptoms combined for cases ≥5 years of age (IID2 study 2008–9)

	Baseline model OR (95%CI)	Baseline + NS-SEC OR (95%CI)
Age group (years)		
5–14	reference	reference
15–24	3.01 (1.66–5.57)	2.70 (1.48–5.02)
25–44	1.02 (0.69–1.50)	0.96 (0.65–1.42)
45–64	0.73 (0.51–1.06)	0.69 (0.47–1.00)
65+	0.64 (0.43–0.95)	0.60 (0.41–0.90)
Sex		
Female	reference	reference
Male	0.95 (0.77–1.19)	0.90 (0.72–1.13)
Ethnicity		
White	reference	reference
Non-White	2.11 (1.18–3.83)	2.03 (1.14–3.70)
NS-SEC		
Managerial/professional		reference
Intermediate		1.21 (0.91–1.61)
Routine/manual		2.18 (1.66–2.87)
Log-likelihood	−1255.1	−1239.3
Deviance	2510.2	2478.6
AIC	2526.2	2498.6
BIC	2566.7	2549.2
Number	1164	1164

AIC Akaike information criterion, *BIC* Bayesian information criterion, *CI* confidence interval, *IID* infectious intestinal disease, *NS-SEC* National Statistics Socioeconomic Classification, *OR* odds ratio

Sensitivity analyses

We undertook several robustness tests. Similar results to those reported were observed when analyses were conducted with recurrent episodes of IID included with clustering at the individual level accounted for using mixed-effects models, and when the boundaries of the symptom severity categories were changed so that there was an equal 12 point severity score difference within each category (data not shown). Results from multiply imputed datasets, and analyses involving cases of all ages and stratified results by child and adult age groups, also confirmed those from the main analyses (Additional file 1), however ethnicity was not associated with symptom severity when analyses were performed using the imputed datasets. Additionally, comparable associations were found when investigating predictive factors for the duration of absence among absentees (Additional file 1). Lastly, the

Table 4 Multivariate models for sickness absence due to IID for cases of school/working age (IID2 study 2008–9)

	Baseline model	Baseline + NS-SEC	Baseline + NS-SEC + Severity
	OR (95%CI)[a]	OR (95%CI)[a]	OR (95%CI)[a]
Age (years)	0.98 (0.98–0.99)	0.98 (0.98–0.99)	0.99 (0.98–1.00)
Sex			
Female	reference	reference	reference
Male	0.95 (0.71–1.28)	0.91 (0.68–1.22)	0.92 (0.67–1.26)
Ethnicity			
White	reference	reference	reference
Non-White	2.66 (1.16–7.22)	2.58 (1.12–7.00)	1.91 (0.80–5.31)
NS-SEC			
Managerial/ professional		reference	reference
Intermediate		1.13 (0.77–1.66)	1.05 (0.70–1.59)
Routine/manual		1.83 (1.26–2.69)	1.38 (0.92–2.07)
Symptom severity			
Mild			reference
Moderate			3.60 (2.54–5.14)
Severe			5.27 (3.54–7.93)
Log-likelihood	−531.0	−525.9	−482.4
Deviance	1062.0	1051.9	964.9
AIC	1070.0	1063.9	980.9
BIC	1088.9	1092.1	1018.5
Number	818	818	818

AIC Akaike information criterion, *BIC* Bayesian information criterion, *CI* confidence interval, *IID* infectious intestinal disease, *NS-SEC* National Statistics Socioeconomic Classification, *OR* odds ratio

[a]Since the absence outcome was common, the odds ratios should not be interpreted as risk ratios

appropriateness of combining cases from the IID2 component studies was supported by analyses indicating the relationships between NS-SEC and the outcomes were not significantly different between the cohort and GP presentation studies (Additional file 1).

Discussion

We analysed data from the largest population-based survey of IID conducted in the UK, and found that IID cases of lower SES compared to high were more likely to experience severe symptoms, and were more likely to be absent from work or school. The association between SES and sickness absence was largely explained by greater symptom severity amongst the more disadvantaged groups.

Our findings are comparable to those of other studies that have analysed measures of IID severity and SES, however these studies are sparse in number, and have tended to focus on children under five years of age. Our findings suggest that the association between SES and IID severity is true for the whole (all age) population, not just for young children. We identified one British study which analysed data from the population-based Avon Longitudinal Study of Pregnancy and Childhood (ALSPAC), to assess predictive factors for the duration of diarrhoeal episodes in children less than six months of age [22]. The authors found that infants living in rented versus mortgaged/owned accommodation (a suggested indicator of SES) had greater odds of experiencing diarrhoea for six or more days. However, this association became non-significant after adjustment for duration of breast feeding, with longer spells of breast feeding providing protection against prolonged diarrhoea.

Whilst very few cases were admitted to hospital in our sample (<1%), our findings are somewhat similar to those of studies conducted in hospital settings. At this severe end of the disease spectrum, one UK-based study found low SES was associated with longer time to discharge for children hospitalised with gastroenteritis in univariate analysis [23]. Similarly, among American children less than five years of age hospitalised with gastroenteritis, those enrolled in Medicaid (a proxy measure for low SES) experienced longer average length of stay, compared to children not enrolled, when no other factors were taken into consideration [24]. In contrast, multivariate analysis revealed that education level and income were not related to length of stay for Canadian children less than five years of age hospitalised with rotavirus gastroenteritis, whereas regularly seeing a physician for a medical condition was associated with longer hospital stays [25].

These findings might suggest that the association between SES and IID severity could be mediated by socially patterned factors that impair immune response, such as lack of breast feeding in infancy and multimorbidity [26, 27], both of which are more prevalent among lower socioeconomic groups [28, 29]. Additional biologically plausible mechanisms which might help to explain a greater burden of severe IID in lower socioeconomic groups, but are as yet to be substantiated in this context, include increased levels of chronic stress, smoking, and nutritional deficiencies, all of which display social gradients and are associated with immune system compromise [30–34]. The potential mediating role of immune suppressing variables on the relationship between SES and symptom severity warrants further investigation.

We found IID cases of lower SES compared to high had greater odds of sickness absence due to IID, and this was largely explained by greater symptom severity amongst cases of lower SES. In a cohort of UK civil servants, age adjusted rates of sickness absence due to gastroenteritis, were over six and four times higher for men and women respectively, in lower employment grades compared to high [12]. Conversely, self-reported sickness absence for

gastroenteritis in a cohort of Dutch employees was unrelated to education level in univariate analysis [13]. These conflicting findings may, in part, be due to the different populations studied, since our age, sex and ethnicity adjusted results for absence were akin to those observed in the UK-based study of civil servants [12]. However, neither study investigated the role of symptom severity, which was identified as an important mediator of the relationship between SES and sickness absence in our analysis.

There are several limitations to this analysis. The validity of our results depended upon the unbiased and accurate self-reporting of symptoms and sickness absence among cases. If those of lower SES perceived their symptoms differently to those of higher SES, which has been observed in studies investigating perceptions of pain across socioeconomic groups [35, 36], our results could be a mere artefact of the severity measurement. Nonetheless, the variables used to derive the symptom severity score in our study were related to the presence and duration of symptoms, which are rather more objective measures of severity compared to, for example, a subjective rating of symptom severity from mild to severe.

There was a large amount of missing data, particularly within the NS-SEC and symptom severity variables (Table 1). Listwise deletion as a method of handling missing data can produce unbiased estimates when data are missing completely at random [37]. However, the odds of whether data were missing or not within the NS-SEC and symptom severity variables, were associated with other variables within the dataset, supporting the idea that missing data were missing at random, rather than missing completely at random. Sensitivity analyses were therefore performed using multiple imputation by chained equations to impute missing data values (Additional file 1). Results from multiply imputed datasets confirmed those from the main analyses, suggesting that any bias resulting from the use of listwise deletion, was minimal. Ethnicity however was not associated with symptom severity when analyses were performed using the imputed datasets.

Cases identified in the IID2 cohort and GP presentation studies were combined for this analysis. Individuals in managerial/professional occupations, those aged 55+ years and those of White ethnicity were over-represented in the cohort study compared to the UK population, and individuals in intermediate and routine/manual occupations and those aged 15–24 years in particular were under-represented [15]. Under-representation of lower socioeconomic groups is commonplace in population-based surveys [38], and could limit the external validity of our findings. Nevertheless, the internal validity of our findings should remain unaffected. It is possible that if non-participation or the design of the studies resulted in the under-representation of cases of lower SES who experienced milder symptoms, we may have overestimated the association between low SES and severe symptoms. However, within the cohort study this is unlikely as cases were captured prospectively. The GP presentation study may have been more prone to selection bias, since cases with more severe symptoms and those of lower SES may be more likely to present to their GP for an episode of IID [3], however as shown in Additional file 1, the relationship between NS-SEC and symptom severity was not significantly different between the cohort and GP presentation studies.

There is the potential for different pathogens to infect people of different SES, for example *Listeria* and norovirus have been associated with low SES in some studies [39, 40]. Unfortunately, we were unable to explore the role of pathogen type on the association between SES and symptom severity because for around 58% of the sampled cases no pathogen was identified [15]. The impact of pathogen type on the association between SES and symptom severity is unknown, however the severity of illness likely depends not only on the infecting pathogen but also on host factors and the dose to which the host is exposed [41]. The relationship between SES, pathogen type and IID symptom severity could be explored using a larger sample of cases, since for the majority a pathogen will not be identified.

Finally, the IID2 study also contained a retrospective telephone survey which gave higher IID incidence estimates compared to the IID2 cohort study [15], however we were unable to repeat our analyses with cases identified in the telephone survey because NS-SEC information was not collected. We were also unable to assess inequalities in sickness absence amongst those providing care for IID cases (caregiver informative was not collected) however this may be an interesting avenue for further research.

Conclusions

Our study sheds new light into an under-researched area and indicates that the consequences of having an IID may be unequally shared across socioeconomic groups. These consequences are potentially serious. Loss of working days due to sickness can have important economic consequences and these are likely to be more severe for more disadvantaged groups who might receive less adequate compensation from their employer. Loss of days from school can affect educational attainment [42], suggesting that the unequal effects of IID could exacerbate educational inequalities. Actions that reduce the risk of acquiring IID are unlikely to sufficiently address these inequalities; public health interventions also need to reduce their unequal consequences. Further research is required to understand the mechanisms explaining greater severity of illness in disadvantaged groups, and to identify ways to minimise the differential impact of IID on sickness absence.

Additional file

Additional file 1: Supplementary material and sensitivity analyses. (PDF 387 kb)

Abbreviations

ALSPAC: Avon Longitudinal Study of Pregnancy and Childhood; GAM: Generalised additive model; GP: General practice; IID: Infectious intestinal disease; NS-SEC: National Statistics Socioeconomic Classification; SES: Socioeconomic status; UK: United Kingdom

Acknowledgements

The authors thank Dr. John Harris and Dr. Peng Yin for their statistical advice.

Funding

This work was supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections [grant number NIHR HPRU 2012–10,038] at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of East Anglia, University of Oxford and the Institute of Food Research. Tanith Rose is based at the University of Liverpool. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or PHE.

Availability of data and materials

The datasets analysed during the current study are available in the UK Data Service repository, https://discover.ukdataservice.ac.uk/catalogue/?sn=7820&type=Data%20catalogue.

Authors' contributions

All authors contributed to the conception and design of the study. TR performed the analyses with guidance from DTR and BB. TR drafted the manuscript which was revised critically by DTR, BB, MV, JH, NL, SOB and MW. All authors approved the final version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The IID2 study was granted ethical approval by the North West Research Ethics Committee (07/MRE08/5). Participants gave written informed consent for their anonymised data to be used for future analyses.

Publisher's Note

Author details

[1]NIHR Health Protection Research Unit in Gastrointestinal Infections, Liverpool, UK. [2]Department of Public Health and Policy, University of Liverpool, Liverpool, UK. [3]National Infection Service, Public Health England, London/Birmingham, UK. [4]Health Economics Research Centre, University of Oxford, Oxford, UK. [5]Department of Public Health and Policy Institute of Psychology, Health and Society, University of Liverpool, Whelan Building, Liverpool L69 3GB, UK.

Received: 14 March 2017 Accepted: 15 June 2017
Published online: 23 June 2017

References

1. Tam CC, Rodrigues LC, Viviani L, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut. 2012;61:69–77.
2. Food Standards Agency. The second study of infectious intestinal disease in the community (IID2 Study). 2016. https://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/b18021 Accessed 8 Nov 16.
3. Tam CC, Rodrigues LC, O'Brien SJ. The study of infectious intestinal disease in England: what risk factors for presentation to general practice tell us about potential for selection bias in case-control studies of reported cases of diarrhoea. Int J Epidemiol. 2003;32:99–105.
4. Teschke K, Bellack N, Shen H, et al. Water and sewage systems, socio-demographics, and duration of residence associated with endemic intestinal infectious diseases: a cohort study. BMC Public Health. 2010;10:767.
5. Pockett RD, Adlard N, Carroll S, Rajoriya F. Paediatric hospital admissions for rotavirus gastroenteritis and infectious gastroenteritis of all causes in England: an analysis of correlation with deprivation. Curr Med Res Opin. 2011;27:777–84.
6. Olowokure B, Hawker J, Weinberg J, Gill N, Sufi F. Deprivation and hospital admission for infectious intestinal diseases. Lancet. 1999;353(9155):807–8.
7. Dennehy PH, Cortese MM, Bégué RE, et al. A case-control study to determine risk factors for hospitalization for rotavirus gastroenteritis in U.S. children. Pediatr Infect Dis J. 2006;25(12):1123–31.
8. Biering-Sørensen S, Søndergaard G, Vitting Andersen K, et al. Time trends in socio-economic factors and risk of hospitalisation with infectious diseases in pre-school children 1985-2004: a Danish register-based study. Paediatr Perinat Epidemiol. 2012;26(3):226–35.
9. Wilking H, Höhle M, Velasco E, et al. Ecological analysis of social risk factors for rotavirus infections in berlin, Germany, 2007-2009. Int J Health Geogr. 2012;11:37.
10. Kristensen TR, Jensen SM, Kreiner S, Mikkelsen S. Socioeconomic status and duration and pattern of sickness absence. A 1-year follow-up study of 2331 hospital employees. BMC Public Health. 2010;10:643.
11. North F, Syme SL, Feeney A, et al. Explaining socioeconomic differences in sickness absence: the Whitehall II study. BMJ. 1993;306(6874):361–6.
12. Feeney A, North F, Head J, Canner R, Marmot M. Socioeconomic and sex differentials in reason for sickness absence from the Whitehall II study. Occup Environ Med. 1998;55:91–8.
13. Mohren DC, Swaen GM, Kant I, van Schayck CP, Galama JM. Fatigue and job stress as predictors for sickness absence during common infections. Int J Behav Med. 2005;12(1):11–20.
14. O'Brien SJ, Rait G, Hunter PR, et al. Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). BMC Med Res Methodol. 2010;10:39.
15. Tam C, Viviani L, Adak B, et al. The second study of infectious intestinal disease in the community (IID2 Study): Final report. Food Standards Agency, 2012. https://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/b18021 Accessed 8 Nov 2016.
16. Office for National Statistics. Pension trends. Basingstoke: Palgrave Macmillan; 2005.
17. Office for National Statistics. The National Statistics Socio-economic Classification (NS-SEC). https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 Accessed 8 Nov 2016.
18. Rose D, Pevalin DJ, O'Reilly K, editors. The National Statistics Socio-Economic Classification: origins, development and use. Basingstoke: Palgrave Macmillan; 2005.
19. Office for National Statistics. Super Output Area (SOA). http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas%2D-soas-/index.html Accessed 8 Nov 2016.
20. Harrell FE, editor. Regression Modeling Strategies. New York: Springer-Verlag; 2001.
21. UCLA: Statistical Consulting Group. R Data Analysis Examples: Ordinal Logistic Regression. https://stats.idre.ucla.edu/r/dae/ordinal-logistic-regression/. Accessed 19 June 2017.
22. Baker D, Taylor H, Henderson J. ALSPAC study team. Inequality in infant morbidity: causes and consequences in England in the 1990s. J Epidemiol Community Health. 1998;52(7):451–8.
23. Conway SP, Phillips RR, Panday S. Admission to hospital with gastroenteritis. Arch Dis Child. 1990;65:579–84.
24. Ma L, El Khoury AC, Itzler RF. The burden of rotavirus hospitalizations among Medicaid and non-Medicaid children younger than 5 years old. Am J Public Health. 2009;99(Suppl 2):S398–404.
25. Ford-Jones EL, Wang E, Petric M, Corey P, Moineddin R, Fearon M. Hospitalization for community-acquired, rotavirus-associated diarrhea: a prospective, longitudinal, population-based study during the seasonal outbreak. Arch Pediatr Adolesc Med. 2000;154(6):578–85.

26. Jackson KM, Nazar AM. Breastfeeding, the immune response, and long-term health. J Am Osteopath Assoc. 2006;106(4):203–7.
27. Castle SC, Uyemura K, Rafi A, Akande O, Makinodan T. Comorbidity is a better predictor of impaired immunity than chronological age in older adults. J Am Geriatr Soc. 2005;53(9):1565–9.
28. Oakley LL, Renfrew MJ, Kurinczuk JJ, Quigley MA. Factors associated with breastfeeding in England: an analysis by primary care trust. BMJ Open. 2013;3(6):e002765.
29. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380(9836):37–43.
30. Cohen S, Doyle WJ, Baum A. Socioeconomic status is associated with stress hormones. Psychosom Med. 2006;68(3):414–20.
31. Segerstrom SC, Miller GE. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. Psychol Bull. 2004; 130(4):601–30.
32. Stämpfli MR, Anderson GP. How cigarette smoke skews immune responses to promote infection, lung disease and cancer. Nat Rev Immunol. 2009; 9(5):377–84.
33. Darmon N, Drewnowski A. Does social class predict diet quality? Am J Clin Nutr. 2008;87(5):1107–17.
34. Lund BM, O'Brien SJ. The occurrence and prevention of foodborne disease in vulnerable people. Foodborne Pathog Dis. 2011;8(9):961–73.
35. Dorner TE, Muckenhuber J, Stronegger WJ, Ràsky E, Gustorff B, Freidl W. The impact of socio-economic status on pain and the perception of disability due to pain. Eur J Pain. 2011;15(1):103–9.
36. Miljković A, Stipčić A, Braš M, et al. Is experimentally induced pain associated with socioeconomic status? Do poor people hurt more? Med Sci Monit. 2014;20:1232–8.
37. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013;64(5):402–6.
38. Lorant V, Demarest S, Miermans PJ, Van Oyen H. Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census. Int J Epidemiol. 2007;36(6):1292–9.
39. Phillips G, Tam CC, Rodrigues LC, Lopman B. Risk factors for symptomatic and asymptomatic norovirus infection in the community. Epidemiology & Infection. 2011;139(11):1676–86.
40. Gillespie IA, Mook P, Little CL, et al. Human listeriosis in England, 2001-2007: association with neighbourhood deprivation. Euro Surveillance. 2010;15(27):7–16.
41. O'Brien SJ, Halder SL. GI epidemiology: infection epidemiology and acute gastrointestinal infections. Aliment Pharmacol Ther. 2007;25(6):669–74.
42. Department for Education. The link between absence and attainment at KS2 and KS4: 2012/13 academic year. 2015. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/412638/The_link_between_absence_and_attainment_at_KS2_and_KS4.pdf Accessed 8 Nov 2016.

References

Adams NL, Byrne L, Smith GA, et al. (2016). Shiga Toxin–Producing Escherichia coli O157, England and Wales, 1983–2012. *Emerging Infectious Diseases*, 22(4):590–7.

Adams NL, Rose TC, Hawker J, et al. (2017) Socioeconomic status and infectious intestinal disease in the community: a longitudinal study (IID2 study). *Eur J Public Health*, ckx091.

Adamson J, Ben-Shlomo Y, Chaturvedi N, et al. (2003) Ethnicity, socio-economic position and gender--do they affect reported health-care seeking behaviour? *Soc Sci Med*, 57(5):895–904.

Adlam S, Perera S, Lake R, et al. (2011) Acute gastrointestinal illness in New Zealand: A community study. *Epidemiology and Infection*, 139(2):302–8.

Agboado G, Peters J, Donkin L (2012) Factors influencing the length of hospital stay among patients resident in Blackpool admitted with COPD: a cross-sectional study. *BMJ Open*, 2:e000869.

Ajene AN, Fischer Walker CL, Black RE (2013) Enteric Pathogens and Reactive Arthritis: A Systematic Review of Campylobacter, Salmonella and Shigella-associated Reactive Arthritis. *J Health Popul Nutr*, 31:299–307.

Allison P (2012) *Why You Probably Need More Imputations Than You Think*. Available at http://statisticalhorizons.com/more-imputations (accessed 21/07/17).

Al-Sakkaf A (2015) Domestic food preparation practices: a review of the reasons for poor home hygiene practices. *Health Promotion International*, 30(3):427–37.

Ansari Z (2007) The Concept and Usefulness of Ambulatory Care Sensitive Conditions as Indicators of Quality and Access to Primary Health Care. *Australian Journal of Primary Health*, 13(3):91–110.

Arena C, Amoros JP, Vaillant V, et al. (2014) Acute diarrhea in adults consulting a general practitioner in France during winter: incidence, clinical characteristics, management and risk factors. *BMC Infectious Diseases*, 14:574.

Armon K, Stephenson T, Gabriel V, et al. (2001) Determining the common medical presenting problems to an accident and emergency department. *Arch Dis Child*, 84(5):390–392.

Asaria M, Ali S, Doran T, et al. (2016) How a universal health system reduces inequalities: lessons from England. *J Epidemiol Community Health*, 70:637–43.

Asthana S, Halliday J (2006) *What Works in Tackling Health Inequalities?* Bristol: The Policy Press.

Aveyard H (2010) *Doing a Literature Review in Health and Social Care* (2nd edition). Maidenhead: Open University Press/McGraw-Hill Education.

Azur MJ, Stuart EA, Frangakis C, et al. (2011) Multiple Imputation by Chained Equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1):40–9.

Baker D, Taylor H, Henderson J (1998) Inequality in infant morbidity: causes and consequences in England in the 1990s. *J Epidemiol Community Health*, 52(7):451–8.

Baker MG, McDonald A, Zhang J, et al. (2013) *Infectious Diseases Attributable to Household Crowding in New Zealand: A systematic review and burden of disease estimate*. Wellington: He Kainga Oranga/Housing and Health Research Programme, University of Otago.

Baker WL, White CM, Cappelleri JC, et al. (2009) Understanding heterogeneity in meta-analysis: the role of meta-regression. *International Journal Of Clinical Practice*, 63:1426–34.

Bambra C (2016) *Health Divides*. Bristol: Policy Press.

Bankart MJG, Baker R, Rashid A, et al. (2011) Characteristics of general practices associated with emergency admission rates to hospital: a cross-sectional study. *Emergency Medicine Journal*, 28:558–563.

Bannister B, Gillespie S, Jones J (2009) *Infection: Microbiology and Management* (3rd edition). Oxford: Wiley-Blackwell.

Barker I, Steventon A, Deeny SR (2017) Association between continuity of care in general practice and hospital admissions for ambulatory care sensitive conditions: cross sectional study of routinely collected, person level data. *BMJ*, 356:j84.

Barnett K, Mercer SW, Norbury M, et al. (2012) Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*, 380(9836):37–43.

Bartley M, Blane D (2016) 'Reflections on the legacy of British health inequalities', in Smith KE, Bambra C, Hill S (eds.) *Health Inequalities: Critical Perspectives*. New York: Oxford University Press: 22–33.

Batz MB, Hoffmann S, Morris JG Jr. (2012) Ranking the disease burden of 14 pathogens in food sources in the United States using attribution data from outbreak investigations and expert elicitation. *J Food Prot*, 75(7):1278–91.

Baum A, Garofalo JP, Yali AM (1999) Socioeconomic Status and Chronic Stress: Does Stress Account for SES Effects on Health?. *Annals of the New York Academy of Sciences*, 896:131–144.

Beale N, Peart C, Kay H, et al. (2010) 'ALSPAC' infant morbidity and council tax band: Doctor consultations are higher in lower bands. *European Journal of Public Health*, 20(4):403–8.

Bellemare S, Hartling L, Wiebe N, et al. (2004) Oral rehydration versus intravenous therapy for treating dehydration due to gastroenteritis in children: a meta-analysis of randomised controlled trials. *BMC Med*, 2:11.

Bemis K, Marcus R, Hadler JL (2014) Socioeconomic status and campylobacteriosis, Connecticut, USA, 1999-2009. *Emerging Infectious Diseases*, 20(7):1240–2.

Berkey CS, Hoaglin DC, Mosteller F, et al. (1995) A random effects regression model for meta-analysis. *Statistics in Medicine*, 14:395-411.

Bettany-Saltikov J (2012) *How to do a Systematic Literature Review in Nursing: A step-by-step guide*. Maidenhead: Open University Press/McGraw-Hill Education.

Biau DJ, Kernéis S, Porcher R (2008) Statistics in Brief: The Importance of Sample Size in the Planning and Interpretation of Medical Research. *Clinical Orthopaedics and Related Research*, 466(9):2282–8.

Bibby P, Brindley P (2013) *Urban and Rural Area Definitions for Policy Purposes in England and Wales: Methodology (v1.0)*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239477/RUC11methodologypaperaug_28_Aug.pdf (accessed 01/09/17).

Biering-Sørensen S, Søndergaard G, Vitting Andersen K, et al. (2012) Time trends in socio-economic factors and risk of hospitalisation with infectious diseases in pre-school children 1985–2004: a Danish register-based study. *Paediatric and Perinatal Epidemiology,* 26:226–235.

Black D, Morris JN, Smith C, et al. (1980) *Inequalities in Health: Report of a Research Working Group (The Black Report).* London: Department of Health and Social Security.

Black RE, Levine MM, Clements ML, et al. (1988) Experimental Campylobacter jejuni Infection in Humans. *The Journal of Infectious Diseases*, 157(3):472–9.

Blane D (2005) 'The life course, the social gradient, and health', in Marmot M, Wilkinson RG (eds.) *Social Determinants of Health* (2nd edition). New York: Oxford University Press: 54–77.

Blaxter L, Hughes C, Tight M (2010) *How to Research* (4th edition). Maidenhead: Open University Press/McGraw-Hill Education.

BMJ Best Practice (2016) *Viral gastroenteritis in children*. Available at: http://bestpractice.bmj.com/best-practice/monograph/794/follow-up/complications.html (accessed 09/03/17).

Bodner TE (2008) What Improves with Increased Missing Data Imputations? *Structural Equation Modeling*, 15(4):651–75.

Borenstein M, Hedges LV, Higgins JPT, et al. (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods*, 1(2):97–111.

Borg ML, Modi A, Tostmann A, et al. (2012) Ongoing outbreak of Shigella flexneri serotype 3a in men who have sex with men in England and Wales, data from 2009-2011. *Euro Surveill*, 17(13):20137.

Borgnolo G, Barbone F, Scornavacca G, et al. (1996) A case-control study of Salmonella gastrointestinal infection in Italian children. *Acta Pædiatrica,* 85:804–808.

Bouhlila DS, Sellaouti F (2013) Multiple imputation using chained equations for missing data in TIMSS: a case study. *Large-scale Assessments in Education*, 1:4.

Bozkurt AI, Özgür S, Özçirpici B (1999) The prevalance of diseases with diarrhoea among children under five years old in different socioeconomic levels. *Tr J of Medical Sciences*, 29:561–5.

Bozkurt AI, Özgür S, Özçirpici B (2003) Association between household conditions and diarrheal diseases among children in Turkey: a cohort study. *Pediatrics International*, 45(4):443–51.

Braeckman T, Theeten H, Lernout T, et al. (2014) Rotavirus vaccination coverage and adherence to recommended age among infants in Flanders (Belgium) in 2012. *Euro Surveill*, 19(20):20806.

Braveman P, Gottlieb L (2014) The Social Determinants of Health: It's Time to Consider the Causes of the Causes. *Public Health Reports*, 129(Suppl 2):19–31.

British Medical Association (2016) *Health in all policies: health, austerity and welfare reform*. London: British Medical Association.

British Medical Association (2017) *Health at a price: Reducing the impact of poverty*. London: British Medical Association.

Britton E, Hales S, Venugopal K, et al. (2010) The impact of climate variability and change on cryptosporidiosis and giardiasis rates in New Zealand. *Journal of Water and Health*, 8(3):561–71.

Brunner E, Marmot M (2005) 'Social organization, stress, and health', in Marmot M, Wilkinson RG (eds.) *Social determinants of health* (2nd edition). New York: Oxford University Press: 6–31.

Burkhart DM (1999) Management of acute gastroenteritis in children. *Am Fam Physician*, 60(9):2555–63.

Burls A (2009) *What is Critical Appraisal?* London: Hayward Medical Communications.

Busby J, Purdy S, Hollingworth W (2017a) Calculating hospital length of stay using the Hospital Episode Statistics; a comparison of methodologies. *BMC Health Services Research*, 17:347.

Busby J, Purdy S, Hollingworth W (2017b) How do population, general practice and hospital factors influence ambulatory care sensitive admissions: a cross sectional study. *BMC Family Practice*, 18:67.

Byrne L, Jenkins C, Launders N, et al. (2015) The epidemiology, microbiology and clinical impact of Shiga toxin-producing Escherichia coli in England, 2009-2012. *Epidemiol Infect,* 143(16):3475–87.

Cacho NT, Lawrence RM (2017) Innate Immunity and Breast Milk. *Frontiers in Immunology*, 8:584.

Carlson MDA, Morrison RS (2009) Study Design, Precision, and Validity in Observational Studies. *Journal of Palliative Medicine*, 12(1):77–82.

Castle SC, Uyemura K, Rafi A, et al. (2005) Comorbidity is a better predictor of impaired immunity than chronological age in older adults. *J Am Geriatr Soc*, 53(9):1565–9.

Cawthraw SA, Lind L, Kaijser B, et a. (2000) Antibodies, directed towards Campylobacter jejuni antigens, in sera from poultry abattoir workers. *Clinical and Experimental Immunology*, 122(1):55–60.

Centers for Disease Control and Prevention (2015) *Show Me the Science - Why Wash Your Hands?* Available at: https://www.cdc.gov/handwashing/why-handwashing.html (accessed 24/04/17).

Chandola T, Bartley M, Sacker A, et al. (2003) Health selection in the Whitehall II study, UK. *Social Science and Medicine*, 56(10):2059–72.

Chandra RK (1997) Nutrition and the immune system: an introduction. *Am J Clin Nutr*, 66(2):460S–463S.

Chien PF, Howie PW (2001) Breast milk and the risk of opportunistic infection in infancy in industrialized and non-industrialized settings. *Adv Nutr Res*, 10:69–104.

Chinen J, Shearer WT (2010) Secondary immunodeficiencies, including HIV infection. *J Allergy Clin Immunol*, 125(2 Suppl 2):S195–203.

Chow CM, Leung AKC, Hon KL (2010) Acute gastroenteritis: from guidelines to real life. *Clin Exp Gastroenterol*, 3:97–112.

Christenson B, Ringner A, Blücher C, et al. (1983) An outbreak of campylobacter enteritis among the staff of a poultry abattoir in Sweden. *Scand J Infect Dis*, 15(2):167–72.

Christensen RHB (2015) *A Tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal Package*. Available at: https://cran.r-project.org/web/packages/ordinal/vignettes/clmm2_tutorial.pdf (accessed 10/07/17).

Cohen S, Doyle WJ, Baum A (2006) Socioeconomic status is associated with stress hormones. *Psychosom Med*, 68(3):414–20.

Cohen SA, Egorov AI, Jagai JS, et al. (2008) The SEEDs of two gastrointestinal diseases: socioeconomic, environmental, and demographic factors related to cryptosporidiosis and giardiasis in Massachusetts. *Environ. Res*, 108:185–191.

Collin PH (ed.) (1998) *Dictionary of Medicine*. Chicago: Fitzroy Dearborn Publishers.

Collins B (2013) Food hygiene, deprivation, types of premises and rates of gastrointestinal illnesses in the UK. *Public Health Nutr*, 18(1):160–6.

Commission on Social Determinants of Health (2008) *Closing the gap in a generation: health equity through action on the social determinants of health: Final Report of the Commission on Social Determinants of Health*. Geneva: World Health Organization.

Consumer Data Research Centre (2017) *Access to Healthy Assets and Hazards (AHAH)*. Available at: https://data.cdrc.ac.uk/dataset/access-to-healthy-assets-and-hazards-ahah (accessed 27/11/16).

Conway SP, Phillips RR, Panday S (1990) Admission to hospital with gastroenteritis. *Archives of Disease in Childhood*, 65:579–84.

Covvey JR, Johnson BF, Elliott V, et al. (2014) An association between socioeconomic deprivation and primary care antibiotic prescribing in Scotland. *J Antimicrob Chemother*, 69(3):835-41.

Crichton J, Hickman M, Campbell R, et al. (2015) Socioeconomic factors and other sources of variation in the prevalence of genital chlamydia infections: A systematic review and meta-analysis. *BMC Public Health*, 15:729.

Cuberes D, Roberts J (2015) *Household location and income: a spatial analysis for British cities*. Sheffield Economics Research Papers Series. Available at: https://www.sheffield.ac.uk/economics/research/serps/articles/2015_022 (accessed 01/09/17).

Curtis S (2004) *Health and Inequality: Geographical Perspectives*. London: Sage Publications.

Dahlgren G, Whitehead M (2007) *Levelling up (part 2): a discussion paper on European strategies for tackling social inequities in health.* Copenhagen: World Health Organisation Regional Office for Europe.

Darmon N, Drewnowski A (2008) Does social class predict diet quality? *Am J Clin Nutr*, 87(5):1107–17.

Darmon N, Drewnowski A (2015) Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. *Nutrition Reviews*, 73(10):643–660.

Davies HTO, Crombie IK, Tavakoli M (1998) When can odds ratios mislead? *BMJ*, 316(7136): 989–91.

Davies R (2017) Sick Parcelforce couriers can be charged up to £250 if they can't find cover. *The Guardian*. Available at: https://www.theguardian.com/business/2017/mar/06/sick-parcelforce-couriers-royal-mail-mands-john-lewis-hamleys-dpd (accessed 14/09/17).

de Perio MA, Niemeier R, Levine SJ, et al. (2013) Campylobacter Infection in Poultry-Processing Workers, Virginia, USA, 2008–2011. *Emerging Infectious Diseases*, 19(2):286–8.

De Wit MA, Koopmans MP, Kortbeek LM, et al. (2001a) Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. *American Journal of Epidemiology*, 154(7):666–74.

De Wit MA, Kortbeek LM, Koopmans MP, et al. (2001b) A comparison of gastroenteritis in a general practice-based study and a community-based study. *Epidemiol Infect*, 127(3):389–97.

Deaton A (2013) 'What does the empirical evidence tell us about the injustice of health inequalities?', in Eyal N, Hurst S, Norheim OF et al. (eds.) *Inequalities in Health: Concepts, Measures, and Ethics*. New York: Oxford University Press: 263–82.

Deeks JJ, Higgins JPT, Altman DG, et al. (2011) 'Chapter 9: Analysing data and undertaking meta-analyses', in Higgins JPT, Green S (eds.) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available at www.handbook.cochrane.org (accessed 26/11/17).

Dennehy PH (2000) Transmission of rotavirus and other enteric pathogens in the home. *Pediatr Infect Dis J*, 19(10 Suppl):S103–5.

Dennehy PH, Cortese MM, Begue RE, et al. (2006) A case-control study to determine risk factors for hospitalization for rotavirus gastroenteritis in U.S. children. *Pediatric Infectious Disease Journal*, 25(12):1123–31.

Department for Communities and Local Government (2015) *The English Indices of Deprivation 2015: Statistical Release*. Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (accessed 01/09/17).

Department for Education (2015) *The link between absence and attainment at KS2 and KS4: 2012/13 academic year*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/412638/The_link_between_absence_and_attainment_at_KS2_and_KS4.pdf (accessed 08/11/16).

Department for Education (2016) *The link between absence and attainment at KS2 and KS4: 2013/14 academic year*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/509679/The-link-between-absence-and-attainment-at-KS2-and-KS4-2013-to-2014-academic-year.pdf (accessed 14/09/17).

Diaz-Caneja A, Gledhill J, Weaver T, et al. (2005) A child's admission to hospital: a qualitative study examining the experiences of parents. *Intensive Care Med*, 31(9):1248–54.

Diderichsen F, Evans T, Whitehead M (2001) 'The social basis of disparities in health', in Evans T, Whitehead M, Diderichsen F, et al. (eds.) *Challenging Inequities in Health: From Ethics to Action*. New York: Oxford University Press: 12–24.

Diderichsen F, Hallqvist J (1998) 'Social inequalities in health: some methodological considerations for the study of social position and social context', in Arve-Parès B (ed.) *Inequality in Health – A Swedish Perspective*. Stockholm: Swedish Council for Social Research: 25–39.

Dieterich CM, Felice JP, O'Sullivan E, et al. (2013) Breastfeeding and Health Outcomes for the Mother-Infant Dyad. *Pediatr Clin North Am*, 60(1):31–48.

Domnich A, Panatto D, Gasparini R, et al. (2012) The "healthy immigrant" effect: does it exist in Europe today? *Italian Journal of Public Health*, 9(3):e7532.

Dong Y, Peng C-YJ (2013) Principled missing data methods for researchers. *SpringerPlus*, 2:222.

Doorduyn Y, Van Pelt W, Havelaar AH (2012) The burden of infectious intestinal disease (IID) in the community: a survey of self-reported IID in The Netherlands. *Epidemiology and Infection*, 140(7):1185–92.

Dorner TE, Muckenhuber J, Stronegger WJ, et al. (2011) The impact of socio-economic status on pain and the perception of disability due to pain. *Eur J Pain*, 15(1):103–9.

Douglas M (2016) 'Beyond "health": why don't we tackle the cause of health inequalities?', in Smith KE, Bambra C, Hill S (eds.) *Health Inequalities: Critical Perspectives*. New York: Oxford University Press: 109–24.

Duijts L, Jaddoe VW, Hofman A, et al. (2010) Prolonged and Exclusive Breastfeeding Reduces the Risk of Infectious Diseases in Infancy. *Pediatrics*, 126(1):e18–25.

Eaton-Evans J, Dugdale AE (1987) Effects of feeding and social factors on diarrhoea and vomiting in infants. *Archives of Disease in Childhood*, 62(5):445–8.

Eckert J (2005) 'Protozoa', in Kayser FH, Bienz KA, Eckert J et al. *Medical Microbiology*. Stuttgart: Georg Thieme Verlag: 476–543.

Egger M, Smith GD, Schneider (2001) 'Systematic reviews of observational studies', in Egger M, Smith GD, Altman DG (eds.) *Systematic reviews in health care: meta-analysis in context* (2nd edition). London: BMJ Publishing Group: 211–227.

Eldredge JD (2000) Evidence-based librarianship: an overview. *Bull Med Libr Assoc*, 88(4): 289–302.

Elliot AJ, Hughes HE, Astbury J, et al. (2016) The potential impact of media reporting in syndromic surveillance: an example using a possible Cryptosporidium exposure in North West England, August to September 2015. *Euro Surveill*, 21(41):30368.

Elliott EJ (2007) Acute gastroenteritis in children. *BMJ*, 334(7583):35–40.

Esan OB, Pearce M, van Hecke O, et al. (2017) Factors Associated with Sequelae of Campylobacter and Non-typhoidal Salmonella Infections: A Systematic Review. *EBioMedicine*, 15:100–111.

Espelt A, Borrell C, Palència L, et al. (2013) Socioeconomic inequalities in the incidence and prevalence of type 2 diabetes mellitus in Europe. *Gac Sanit*, 27(6):494–501.

Etiler N, Velipasaoglu S, Aktekin M (2004) Risk factors for overall and persistent diarrhoea in infancy in Antalya, Turkey: A cohort study. *Public Health*, 118(1):62–9.

Evans GW, Kim P (2013) Childhood Poverty, Chronic Stress, Self-Regulation, and Coping. *Child Dev Perspect*, 7:43–48.

Evans GW, Schamberg MA (2009) Childhood poverty, chronic stress, and adult working memory. *Proc Natl Acad Sci USA*, 106(16):6545–9.

Evans MR, Sarvotham T, Thomas DR, et al. (2006) Domestic and travel-related foodborne gastrointestinal illness in a population health survey. *Epidemiol Infect*, 134(4):686–93.

Farthing M, Feldman R, Finch R, et al. (1996) The management of infective gastroenteritis in adults. A consensus statement by an expert panel convened by the British Society for the Study of Infection. *Journal of Infection*, 33(3):143–52.

Farthing M, Salam M, Lindberg G, et al. (2012) *Acute diarrhea in adults and children: a global perspective*. Milwaukee: World Gastroenterology Organisation.

Fauci AS (2001) Infectious diseases: considerations for the 21st century. *Clin Infect Dis*, 32:675–85.

Feeney A, North F, Head J, et al. (1998) Socioeconomic and sex differentials in reason for sickness absence from the Whitehall II study. *Occup Environ Med*, 55:91–8.

Feigin VL, Lawes CM, Bennett DA, et al. (2009) Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review. *Lancet Neurol*, 8(4):355–69.

Fein SB, Lin CTJ, Levy AS (1995) Foodborne illness: Perceptions, experience, and preventive behaviors in the United States. *Journal of Food Protection*, 58(12):1405–11.

Field A, Miles J, Field Z (2012) *Discovering Statistics Using R*. London: Sage.

Field F, Allen H, Black M, et al. (2017) *Self-employment and the gig economy: Thirteenth Report of Session 2016–17*. Available at: https://publications.parliament.uk/pa/cm201617/cmselect/cmworpen/847/847.pdf (accessed 14/09/17).

Fink A (2010) *Conducting Research Literature Reviews: From the Internet to Paper* (3rd edition). Thousand Oaks: Sage.

Fisher A, Dembry LM (2017) Norovirus and Clostridium difficile outbreaks: squelching the wildfire. *Curr Opin Infect Dis*, 30(4):440–7.

Flacking R, Hedberg Nyqvist K, Ewald U (2007) Effects of socioeconomic status on breastfeeding duration in mothers of preterm and term infants. *European Journal of Public Health*, 17(6):579–84.

Fletcher SM, McLaws ML, Ellis JT (2013) Prevalence of Gastrointestinal Pathogens In Developed and Developing Countries: Systematic Review and Meta-Analysis. *Journal of Public Health Research*, 2(1):42–53.

Food Standards Agency (2011) *Foodborne Disease Strategy 2010-15: an FSA programme for the reduction of foodborne disease in the UK*. Available at: https://www.food.gov.uk/sites/default/files/multimedia/pdfs/fds2015.pdf (accessed 25/09/17).

Food Standards Agency (2012) *Annual Report of the Chief Scientist 2011/12*. Available at: https://www.food.gov.uk/sites/default/files/multimedia/pdfs/publication/csar1112.pdf (accessed 24/04/17).

Food Standards Agency (2016) *The second study of infectious intestinal disease in the community (IID2 Study)*. Available at: https://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/b18021 (accessed 24/04/17).

Food Standards Agency (no date) *Campylobacter*. Available at: https://www.food.gov.uk/science/microbiology/campylobacterevidenceprogramme (accessed 24/04/17).

Ford-Jones EL, Wang E, Petric M, et al. (2000) Hospitalization for community-acquired, rotavirus-associated diarrhea: a prospective, longitudinal, population-based study during the seasonal outbreak. *Arch Pediatr Adolesc Med*, 154(6):578–85.

Fox J, Weisberg S (2011) *An R Companion to Applied Regression* (2nd edition). Thousand Oaks: Sage.

Galama TJ, van Kippersluis H (2013) 'Health inequalities through the lens of health-capital theory: issues, solutions, and future directions', in Dias PR, O'Donnell O (eds.) *Health and Inequality*. Bingley: Emerald Group Publishing Limited: 263–85.

Gardner W, Mulvey EP, Shaw EC (1995) Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin*, 118:392–404.

Garg AX, Hackam D, Tonelli M (2008) Systematic Review and Meta-analysis: When One Study Is Just not Enough. *Clin J Am Soc Nephrol*, 3(1):253–60.

Gartner A, Farewell D, Dunstan F, et al. (2008) Differences in mortality between rural and urban areas in England and Wales, 2002-04. *Health Stat Q*, (39):6–13.

Giannattasio A, Guarino A, Lo Vecchio A (2016) Management of children with prolonged diarrhea. *F1000Research*, 5:F1000 Faculty Rev-206.

Gibney KB, O'Toole J, Sinclair M, et al. (2014) Disease burden of selected gastrointestinal pathogens in Australia, 2010. *International Journal of Infectious Diseases*, 28:176–185.

Gibson OR, Segal L, McDermott RA (2013) A systematic review of evidence on the association between hospitalisation for chronic disease related ambulatory care sensitive conditions and primary health care resourcing. *BMC Health Services Research*, 13:336.

Giesinger I, Goldblatt P, Howden-Chapman P, et al. (2014) Association of socioeconomic position with smoking and mortality: The contribution of early life circumstances in the 1946 birth cohort. *Journal of Epidemiology and Community Health*, 68(3):275–9.

Gilens M (2005) Inequality and Democratic Responsiveness. *Public Opinion Quarterly*, 69(5):778–96.

Gillespie IA, Mook P, Little CL, et al. (2010) Human listeriosis in England, 2001-2007: association with neighbourhood deprivation. *Euro Surveill*, 15:7–16.

Glass RI, Parashar UD, Bresee JS, et al. (2006) Rotavirus vaccines: current prospects and future challenges. *Lancet*, 368(9532):323–32.

Goldman D (2000) Chronic lymphocytic leukemia and its impact on the immune system. *Clin J Oncol Nurs*, 4(5):233–4.

Gough D, Oliver S, Thomas J (2012) 'Introducing systematic reviews', in Gough D, Oliver S, Thomas J (eds.) *An introduction to systematic reviews*. London: Sage Publications: 1–16.

GOV.UK (2010) *Notifiable diseases and causative organisms: how to report*. Available at: https://www.gov.uk/guidance/notifiable-diseases-and-causative-organisms-how-to-report (accessed 24/04/17).

GOV.UK (2013) *Rotavirus vaccination programme for infants*. Available at: https://www.gov.uk/government/collections/rotavirus-vaccination-progarmme-for-infants (accessed 25/09/17).

GOV.UK (2014) *Notifications of infectious diseases (NOIDs)*. Available at: https://www.gov.uk/government/collections/notifications-of-infectious-diseases-noids (accessed 24/04/17).

Graham H (2004) Social determinants and their unequal distribution: clarifying policy understandings. *Milbank Q*, 82(1):101–24.

Graham H (2007) *Unequal Lives: Health and Socio-Economic Inequalities*. Maidenhead: Open University Press/McGraw-Hill Education.

Graham H (2009) 'Introduction: the challenge of health inequalities', in Graham H (ed.) *Understanding Health Inequalities* (2nd edition). Maidenhead: Open University Press/McGraw-Hill Education: 1–25.

Graham JW (2009) Missing data analysis: making it work in the real world. *Annual Review of Psychology*, 60:549–76.

Graham JW, Olchowski AE, Gilreath TD (2007) How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science,* 8(3):206–13.

Gray AM (1982) Inequalities in Health. The Black Report: A Summary and Comment. *Int J Health Serv*, 12(3):349–80.

Green MJ, Leyland AH, Sweeting H, et al. (2016) Socioeconomic position and early adolescent smoking development: evidence from the British Youth Panel Survey (1994–2008). *Tobacco Control*, 25:203–10.

Greenhalgh T (2010) *How to Read a Paper: The basics of evidence-based medicine* (4th edition). Chichester: Wiley-Blackwell/BMJ Books.

Greenhalgh T, Peacock R (2005) Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*, 331 (7524):1064–5.

Greenland S, Robins J (1994) Ecologic studies – biases, misconceptions, and counterexamples. *Am J Epidemiol*, 139:747–60.

Grootendorst PV (2002) 'A comparison of alternative models of prescription drug utilisation', in Jones A, O'Donnel O (eds.) *Econometric Analysis of Health Data*. Chichester: John Wiley & Sons: 73–87.

Guarino A, Immacolata Spagnuolo M, Russo S, et al. (1995) Etiology and Risk Factors of Severe and Protracted Diarrhea. *Journal of Pediatric Gastroenterology*, 20:173–178.

Hall AJ, Curns AT, McDonald LC, et al. (2012) The Roles of Clostridium difficile and Norovirus Among Gastroenteritis-Associated Deaths in the United States, 1999–2007. *Clin Infect Dis*, 55(2):216–23.

Ham C, Imison C, Jennings M (2010) *Avoiding Hospital Admissions: Lessons from evidence and experience*. London: The King's Fund.

Hamer D, Gorbach S (2010) 'Gastrointestinal Infections', in Warrell D, Cox T, Firth J (eds.) *Oxford textbook of medicine* (5th edition). Oxford: Oxford University Press: 2424–35.

Hannu T (2011) Reactive arthritis. *Best Pract Res Clin Rheumatol*, 25(3):347–57.

Hanson LA, Korotkova M, Håversen L, et al. (2002) Breast-feeding, a complex support system for the offspring. *Pediatr Int*, 44(4):347–52.

Harrell FE (2001) *Regression Modeling Strategies*. New York: Springer-Verlag.

Harrison MJ, Dusheiko M, Sutton M, et al. (2014) Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. *BMJ*, 349:g6423.

Hart C (1998) *Doing a Literature Review: Releasing the Social Science Research Imagination*. London: Sage.

Hart C, Davey Smith G, Blane D (1998) Social mobility and 21 year mortality in a cohort of Scottish men. *Social Science and Medicine*, 47(8):1121–30.

Hart JT (1971) The inverse care law. *The Lancet*, 297:405–12.

Hartling L, Bellemare S, Wiebe N, et al. (2006) Oral versus intravenous rehydration for treating dehydration due to gastroenteritis in children. *Cochrane Database Syst Rev*, (3):CD004390.

Hastie T, Tibshirani R (1990) *Generalized Additive Models*. London: Chapman and Hall.

Havelaar AH, Haagsma JA, Mangen MJ, et al. (2012) Disease burden of foodborne pathogens in the Netherlands, 2009. *Int J Food Microbiol*, 156(3):231–8.

Hawker J, Begg N, Blair I, et al. (2008) *Communicable Disease Control Handbook* (2nd edition). Oxford: Blackwell Publishing.

Health and Social Care Information Centre (2014) *Methodology to create provider and CIP spells from HES APC data*. Available at: https://indicators.hscic.gov.uk/download/Additional%20Reading/Methods%20annexes/Compendium%20User%20Guide%202015%20Feb%20Annex%2013%20V1.pdf (accessed 16/01/17).

Health and Social Care Information Centre (2015) *Hospital Episode Statistics (HES) Analysis Guide*. Available at: http://content.digital.nhs.uk/media/1592/HES-analysis-guide/pdf/HES_Analysis_Guide_Jan_2014.pdf (assessed 03/10/17).

Health Protection Agency (2011) *Gastrointestinal infections annual report, 2010*. London: Health Protection Agency.

Heck KE, Braveman P, Cubbin C, et al. (2006) Socioeconomic Status and Breastfeeding Initiation Among California Mothers. *Public Health Reports*, 121(1):51–9.

Hempel S, Miles JN, Booth MJ, et al. (2013) Risk of bias: a simulation study of power to detect study-level moderator effects in meta-analysis. *Systematic Reviews*, 2:107.

Herikstad H, Yang S, Van Gilder TJ, et al. (2002) A population-based estimate of the burden of diarrhoeal illness in the United States: FoodNet, 1996-7. *Epidemiology and Infection*, 129(1):9–17.

Heys M, Rajan M, Blair M (2017) Length of paediatric inpatient stay, socio-economic status and hospital configuration: a retrospective cohort study. *BMC Health Services Research*, 17:274.

Hicks CM (2009) *Research Methods for Clinical Therapists: Applied Project Design and Analysis* (5th edition). Edinburgh: Churchill Livingstone Elsevier.

Higgins JPT, Altman DG, Sterne JAC, et al. (2011) 'Chapter 8: Assessing risk of bias in included studies', in Higgins JPT, Green S (eds.) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available at www.handbook.cochrane.org (accessed 26/11/17).

Higgins JPT, Deeks JJ (eds.) (2011) 'Chapter 7: Selecting studies and collecting data', in Higgins JPT, Green S (eds.) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available at www.handbook.cochrane.org (accessed 26/11/17).

Hill AB (1965) The Environment and Disease: Association or Causation? *Proceedings of the Royal Society of Medicine*, 58(5):295–300.

Hizarcioglu-Gulsen H, Saltik-Temizel IN, Demir H, et al. (2014) Intractable diarrhea of infancy: 10 years of experience. *J Pediatr Gastroenterol Nutr*, 59(5):571–6.

Hosseinpoor AR, Bergen N, Mendis S, et al. (2012) Socioeconomic inequality in the prevalence of noncommunicable diseases in low- and middle-income countries: Results from the World Health Survey. *BMC Public Health*, 12:474.

Howie PW, Forsyth JS, Ogston SA, et al. (1990) Protective effect of breast feeding against infection. *BMJ*, 300(6716):11–16.

Hsu DC, Katelaris CH (2009) Long-term management of patients taking immunosuppressive drugs. *Aust Prescr*, 32:68–71.

Hungerford D, Vivancos R, French N, et al. (2014) Ecological assessment of the direct and indirect effects of routine rotavirus vaccination in Merseyside, UK using data from multiple health systems: a study protocol. *BMJ Open*, 4:e006161.

Huppertz H, Salman N, Giaquinto C (2008) Risk Factors for Severe Rotavirus Gastroenteritis. *Pediatr Infect Dis J*, 27(1):S11–S19.

Information Standards Board for Health and Social Care (2013) *Anonymisation Standard for Publishing Health and Social Care Data*. Available at: http://content.digital.nhs.uk/isce/publication/isb1523 (assessed 03/10/17).

Irwin A, Scali E (2010) *Action on the Social Determinants of Health: learning from previous experiences. Social Determinants of Health Discussion Paper 1*. Geneva: World Health Organization.

Iturriza Gómara M, Simpson R, Perault AM, et al. (2008) Structured surveillance of infantile gastroenteritis in East Anglia, UK: incidence of infection with common viral gastroenteric pathogens. *Epidemiol Infect*, 136(1):23–33.

Iturriza-Gómara M, Lopman B (2014) Norovirus in healthcare settings. *Curr Opin Infect Dis*, 27(5):437–43.

Jackson KM, Nazar AM (2006) Breastfeeding, the Immune Response, and Long-term Health. *J Am Osteopath Assoc*, 106(4):203–7.

Janssen R, Krogfelt KA, Cawthraw SA, et al. (2008) Host-Pathogen Interactions in Campylobacter Infections: the Host Perspective. *Clin Microbiol Rev*, 21(3):505–18.

Jayaweera H (2014) *Health of Migrants in the UK: What Do We Know?* Migration Observatory Briefing, COMPAS, University of Oxford. Available at: http://www.migrationobservatory.ox.ac.uk/resources/briefings/health-of-migrants-in-the-uk-what-do-we-know/ (accessed 14/04/18).

Johansson G, Lundberg I (2009) Components of the illness flexibility model as explanations of socioeconomic differences in sickness absence. *Int J Health Serv*, 39(1):123–38.

Johnson PCD, Barry SJE, Ferguson HM, et al. (2015) Power analysis for generalized linear mixed models in ecology and evolution. *Methods Ecol Evol*, 6:133–42.

Kaikkonen R, Härkänen T, Rahkonen O, et al. (2015) Explaining educational differences in sickness absence: a population-based follow-up study. *Scand J Work Environ Health*, 41(4):338–46.

Kane R, Wellings K, Free C, et al. (2000) Uses of routine datasets in the evaluation of health promotion interventions: opportunities and limitations. *Health Education*, 100(1):33–41.

Kang H (2013)The prevention and handling of the missing data. *Korean J Anesthesiol*, 64(5):402–6.

Kangovi S, Barg FK, Carter T, et al. (2013) Understanding Why Patients Of Low Socioeconomic Status Prefer Hospitals Over Ambulatory Care. *Health Aff*, 32(7):1196–203.

Kaplan GA, Pamuk ER, Lynch JW, et al. (1996) Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *Br Med J*, 312:999–1003.

Karafillakis E, Hassounah S, Atchison C (2015) Effectiveness and impact of rotavirus vaccines in Europe, 2006-2014. *Vaccine*, 33(18):2097–107.

Katikireddi SV, Higgins M, Smith KE, et al. (2013) Editorial: Health inequalities – the need to move beyond bad behaviours. *Journal of Epidemiology and Community Health*, 67(9):715–6.

Kavanagh D, Richards A, Atkinson J (2008) Complement regulatory genes and hemolytic uremic syndromes. *Annu Rev Med*, 59:293–309.

Keithlin J, Sargeant J, Thomas MK, et al. (2014) Systematic review and meta-analysis of the proportion of Campylobacter cases that develop chronic sequelae. *BMC Public Health*, 14:1203.

Khan K, Kunz R, Kleijnen J, et al. (2011) *Systematic Reviews to Support Evidence-Based Medicine* (2nd edition). London: Hodder Arnold.

Khan KS, Kunz R, Kleijnen J, et al. (2003) Five steps to conducting a systematic review. *J R Soc Med*, 96(3):118–21.

Kheifets LI, Afifi AA, Buffler PA, et al. (1995) Occupational Electric and Magnetic Field Exposure and Brain Cancer: A Meta-Analysis. *J Occup Environ Med*, 37(12):1327–41.

Kim S, Egerter S, Cubbin C, et al. (2007) Potential Implications of Missing Income Data in Population-Based Surveys: An Example from a Postpartum Survey in California. *Public Health Reports*, 122(6):753–63.

Koller D, Hoffmann F, Maier W, et al. (2013) Variation in antibiotic prescriptions: is area deprivation an explanation? Analysis of 1.2 million children in Germany. *Infection*, 41:121–7.

Kristensen TR, Jensen SM, Kreiner S, et al. (2010) Socioeconomic status and duration and pattern of sickness absence. A 1-year follow-up study of 2331 hospital employees. *BMC Public Health*, 10:643.

Kristenson M (2006) 'Socio-economic position and health: the role of coping', in Siegrist J, Marmot M (eds.) *Social Inequalities in Health: New evidence and policy implications*. New York: Oxford University Press: 127–53.

Kum-Nji P, Mangrem CL, Wells PJ, et al. (2009) Is environmental tobacco smoke exposure a risk factor for acute gastroenteritis in young children? *Clinical Pediatrics*, 48(7):756–62.

Kurts C, Panzer U, Anders HJ, et al. (2013) The immune system and kidney disease: basic concepts and clinical implications. *Nature Reviews Immunology*, 13:738–53.

Kyle RG, Kukanova M, Campbell M, et al. (2011) Childhood disadvantage and emergency admission rates for common presentations in London: an exploratory analysis. *Archives of Disease in Childhood*, 96(3):221–6.

Lake RJ, Cressey PJ, Campbell DM, et al. (2010) Risk ranking for foodborne microbial hazards in New Zealand: burden of disease estimates. *Risk Anal*, 30(5):743–52.

Lal A, Baker MG, French NP, et al. (2012) The epidemiology of human salmonellosis in New Zealand, 1997-2008. *Epidemiology and Infection*, 140(9):1685–94.

LaMorte WW (2016a) *Risk Ratios and Rate Ratios (Relative Risk)*. Available at: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_Association/EP713_Association3.html (accessed 26/11/17).

LaMorte WW (2016b) *Confidence Intervals for Measures of Association*. Available at: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_RandomError/EP713_RandomError4.html (accessed 26/11/17).

LaMorte WW (2017) *Confidence Interval for an Odds Ratio*. Available at: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/QuantCore/PH717_ComparingFrequencies/PH717_ComparingFrequencies8.html (accessed 26/11/17).

Lamps LW (2009) *Surgical pathology of the gastrointestinal system: bacterial, fungal, viral and parasitic infections*. New York: Springer.

Langdon A, Crook N, Dantas G (2016) The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Medicine*, 8:39.

Lantz PM, House JS, Mero RP, et al. (2005) Stress, life events, and socioeconomic disparities in health: results from the Americans' Changing Lives Study. *J Health Soc Behav*, 46(3):274–88.

LaRocque RC, Calderwood SB (2015) 'Syndromes of enteric infection', in Bennett JE, Dolin R, Blaser MJ (eds.) *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases* (8th edition). Philadelphia: Elsevier Saunders: 1238–48.

Lee J, Tan CS, Chia KS (2009) A Practical Guide for Multivariate Analysis of Dichotomous Outcomes. *Annals Academy of Medicine*, 38(8):714–9.

Lefebvre C, Manheimer E, Glanville J, et al. (2011) 'Chapter 6: Searching for studies', in Higgins JPT, Green S (eds.) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available at www.handbook.cochrane.org (accessed 26/11/17).

Leggett HC, Cornwallis CK, West SA (2012) Mechanisms of Pathogenesis, Infective Dose and Virulence in Human Parasites. *PLoS Pathog*, 8(2):e1002512.

Lim JY, Yoon JW, Hovde CJ (2010) A Brief Overview of Escherichia coli O157:H7 and Its Plasmid O157. *Journal of Microbiology and Biotechnology*, 20(1):5–14.

Link BG, Phelan J (1995) Social conditions as fundamental causes of disease. *J Health Soc Behav*, (Extra Issue): 80–94.

Little RJA, Rubin DB (2002) *Statistical Analysis with Missing Data* (2nd edition). Hoboken: John Wiley and Sons.

Lorant V, Demarest S, Miermans PJ, et al. (2007) Survey error in measuring socio-economic risk factors of health status: a comparison of a survey and a census. *Int J Epidemiol*, 36(6):1292–9.

Lorgelly PK, Joshi D, Iturriza Gómara M, et al. (2008) Infantile gastroenteritis in the community: a cost-of-illness study. *Epidemiology and Infection*, 136(1):34–43.

Ludvigsson JF, Abis Study Group, et al. (2006) Epidemiological study of constipation and other gastrointestinal symptoms in 8000 children. *Acta Paediatrica*, 95(5):573–80.

Lund BM, O'Brien SJ (2011) The Occurrence and Prevention of Foodborne Disease in Vulnerable People. *Foodborne Pathog Dis*, 8(9):961–73.

Ma L, El Khoury AC, Itzler RF (2009) The Burden of Rotavirus Hospitalizations Among Medicaid and Non-Medicaid Children Younger Than 5 Years Old. *Am J Public Health*, 99(Suppl 2):S398–S404.

MacIntyre S (1997) The Black Report and beyond: What are the issues? *Social Science and Medicine*, 44(6):723–45.

MacIntyre S, Ellaway A (2009) 'Neighbourhood influences on health', in Graham H (ed.) *Understanding Health Inequalities* (2nd edition). Maidenhead: Open University Press/McGraw-Hill Education: 84–100.

Mackenbach JP, Bakker MJ, Kunst AE, et al. (2002) 'Socioconomic inequalities in health in Europe: An overview', in Mackenbach JP, Bakker MJ (eds.) *Reducing inequalities in health: a European perspective*. London: Routledge: 3–25.

Mackenbach JP (2012) The persistence of health inequalities in modern welfare states: the explanation of a paradox. *Social Science & Medicine*, 75:761–9.

Mackinnon DP (2011) Integrating Mediators and Moderators in Research Design. *Res Soc Work Pract*, 21: 675–681.

Macleod M (2004) Inequalities in the incidence and management of, and outcome from, neurological diseases in older people. *Reviews in Clinical Gerontology*, 13:1–8.

Magán P, Alberquilla A, Otero A, et al. (2011) Hospitalizations for ambulatory care sensitive conditions and quality of primary care: their relation with socioeconomic and health care variables in the Madrid regional health service (Spain). *Med Care*, 49(1):17–23.

Maguire ER, Monsivais P (2015) Socio-economic dietary inequalities in UK adults: an updated picture of key food groups and nutrients from national surveillance data. *The British Journal of Nutrition*, 113(1):181–189.

Majowicz SE, Doré K, Flint JA, et al. (2004) Magnitude and distribution of acute, self-reported gastrointestinal illness in a Canadian community. *Epidemiol Infect*, 132(4):607–17.

Majowicz SE, Horrocks J, Bocking K (2007) Demographic determinants of acute gastrointestinal illness in Canada: a population study. *BMC Public Health*, 7:162.

Malek MH, Berger DE, Coburn JW (2007) On the inappropriateness of stepwise regression analysis for model building and testing. *Eur J Appl Physiol*, 101(2):263–4.

Mandela N (2005) *Make Poverty History*. 3 February, Trafalgar Square, London.

Mangrio E, Wremp A, Moghaddassi M, et al. (2009) Antibiotic use among 8-month old children in Malmö, Sweden in relation to child characteristics and parental sociodemographic, psychosocial and lifestyle factors. *BMC Pediatrics*, 9:31.

Marmot M (2005) 'Introduction', in Marmot M, Wilkinson RG (eds.) *Social determinants of health* (2nd edition). New York: Oxford University Press: 1–6.

Marmot M (2013) 'Fair society healthy lives', in Eyal N, Hurst S, Norheim OF et al. (eds.) *Inequalities in Health: Concepts, Measures, and Ethics*. New York: Oxford University Press: 282–99.

Marmot M, Atkinson T, Bell J, et al. (2010) *Fair Society, Healthy Lives: The Marmot Review. Strategic Review of Health Inequalities in England Post-2010*. London: University College London.

Marmot M, Bosma H, Hemingway H, et al. (1997) Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet*, 350(9073): 235–9.

Marmot M, Shipley MJ, Rose G (1984) Inequalities in death – specific explanations of a general pattern? *Lancet*, 323(8384):1003–6.

Marshall A, Altman DG, Royston P, et al. (2010) Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol*, 10:7.

Marshall JK, Thabane M, Borgaonkar MR, et al. (2007) Postinfectious irritable bowel syndrome after a food-borne outbreak of acute gastroenteritis attributed to a viral pathogen. *Clin Gastroenterol Hepatol*, 5(4):457–60.

McCullagh P (1980) Regression Models for Ordinal Data. *Journal of the Royal Statistical Society*, Series B (Methodological) 42(2):109–42.

McDonagh M, Peterson K, Raina P, et al. (2013) 'Avoiding Bias in Selecting Studies', in *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* [Internet]. Rockville: Agency for Healthcare Research and Quality. Available at: https://www.ncbi.nlm.nih.gov/books/NBK126701/ (accessed 26/11/17).

McGarrol S (2017) *Conversation with Tanith Rose*, 26 September.

McGinnis JM, Williams-Russo P, Knickman JR (2002) The Case For More Active Policy Attention To Health Promotion. *Health Aff*, 21(2):78–93.

McKee M, Pomerleau J (2005) 'Inequalities in health', in Pomerleau J, McKee M (eds.) *Issues in Public Health*. Maidenhead: Open University Press: 80–105.

McKeown T, Brown RG, Record RG (1972) An Interpretation of the Modern Rise of Population in Europe. *Population Studies*, 26(3):345–82.

Mehta H, Nazzal K, Sadikot RT (2008) Cigarette smoking and innate immunity. *Inflammation Research*, 57(11):497–503.

Melzer D, McWilliams B, Brayne C, et al. (2000) Socioeconomic status and the expectation of disability in old age: estimates for England. *Journal of Epidemiology & Community Health*, 54:286-92.

Met Office (2015) *Climate Zones*. Available at: http://www.metoffice.gov.uk/climate-guide/climate/zones (accessed 29/11/16).

Miljković A, Stipčić A, Braš M, et al. (2014) Is experimentally induced pain associated with socioeconomic status? Do poor people hurt more? *Med Sci Monit*, 20:1232–8.

Miller G, Dunn GM, Reid TM, et al. (2005) Does age acquired immunity confer selective protection to common serotypes of Campylobacter jejuni? *BMC Infectious Diseases*, 5:66.

Miller R, Spiro A, Stanner S (2016) Micronutrient status and intake in the UK – where might we be in 10 years' time? *Nutr Bull*, 41:14–41.

Mohren DC, Swaen GM, Kant I, et al. (2005) Fatigue and job stress as predictors for sickness absence during common infections. *Int J Behav Med*, 12(1):11–20.

Moons KG, Donders RA, Stijnen T, et al. (2006) Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*, 59(10):1092–101.

Moorin RE, Heyworth JS, Forbes GM, et al. (2010) Long-term health risks for children and young adults after infective gastroenteritis. *Emerging Infectious Diseases*, 16(9):1440–47.

Morgan A (2006) 'Determinants of Health', in Davies M, Macdowall W (eds.) *Health Promotion Theory.* Maidenhead: Oxford University Press: 49–63.

Nakai M, Weiming K (2011) Review of Methods for Handling Missing Data in Longitudinal Data Analysis. *Int Journal of Math Analysis*, 5(1):1–13.

National Collaborating Centre for Women's and Children's Health (2009) *Diarrhoea and vomiting caused by gastroenteritis: diagnosis, assessment and management in children younger than 5 years.* London: RCOG Press.

National Institute for Health and Care Excellence (2009) *Diarrhoea and vomiting caused by gastroenteritis in under 5s: diagnosis and management*. Available at: https://www.nice.org.uk/guidance/cg84/chapter/1-Guidance (accessed 11/11/17).

National Institute for Health and Care Excellence (2015) *Gastroenteritis*. Available at: https://cks.nice.org.uk/gastroenteritis#!diagnosissub:1 (accessed 19/01/17).

National Institutes of Health (2016) *Health Economics Information Resources: A Self-Study Course: Module 2 - Sources and Characteristics of Information Relating to Health Care Financing in the US*. Available at: https://www.nlm.nih.gov/nichsr/edu/healthecon/02_he_01.html (accessed 27/04/17).

Newburg DS, Ruiz-Palacios GM, Morrow AL (2005) Human milk glycans protect infants against enteric pathogens. *Annu Rev Nutr*, 25:37–58.

Newman KL, Leon JS, Rebolledo PA, et al. (2015) The impact of socioeconomic status on foodborne illness in high-income countries: a systematic review. *Epidemiol Infect*, 143(12):2473–85.

NHS Choices (2016) *Diarrhoea and vomiting (gastroenteritis).* Available at: http://www.nhs.uk/conditions/gastroenteritis/Pages/Introduction.aspx (accessed 19/01/17).

NHS Digital (2016a) *Hospital Admitted Patient Care Activity, 2015-16*. Available at: http://www.content.digital.nhs.uk/catalogue/PUB22378 (accessed 26/04/17).

NHS Digital (2016b) *The HES Processing Cycle and Data Quality.* Available at: http://content.digital.nhs.uk/media/1366/The-HES-processing-cycle-and-HES-data-quality/pdf/HESDQ_In_001_The_HES_Processing_Cycle_and_HES_Data_Quality.pdf (assessed 16/01/17).

NHS Digital (no date) *Hospital Episode Statistics*. Available at: http://content.digital.nhs.uk/hes (assessed 16/01/17).

NHS England (2014) *Emergency admissions for Ambulatory Care Sensitive Conditions – characteristics and trends at national level*. Available at: https://www.england.nhs.uk/wp-content/uploads/2014/03/red-acsc-em-admissions-2.pdf (accessed 01/09/17).

Niedhammer I, Lesuffleur T, Memmi S, et al. (2017) Working conditions in the explanation of occupational inequalities in sickness absence in the French SUMER study. *European Journal of Public Health*, ckx052.

North F, Syme SL, Feeney A, et al. (1993) Explaining socioeconomic differences in sickness absence: the Whitehall II Study. *BMJ*, 306(6874):361–6.

O'Brien SJ (2015) *Conversation with Tanith Rose*.

O'Brien SJ, Halder SLS (2007) GI Epidemiology: infection epidemiology and acute gastrointestinal infections. *Aliment Pharmacol Ther*, 25(6):669–74.

O'Brien SJ, Larose TL, Adak GK, et al. (2016) Modelling study to estimate the health burden of foodborne diseases: cases, general practice consultations and hospitalisations in the UK, 2009. *BMJ Open*, 6(9):e011119.

O'Brien SJ, Rait G, Hunter PR, et al. (2010) Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). *BMC Med Res Methodol*, 10:39.

Oakley LL, Renfrew MJ, Kurinczuk JJ, et al. (2013) Factors associated with breastfeeding in England: an analysis by primary care trust. *BMJ Open*, 3(6).

Office for National Statistics (2005) *Pension Trends.* Basingstoke: Palgrave Macmillan.

Office for National Statistics (2012) *Statistical disclosure control for 2011 Census*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23 (accessed 12/09/17).

Office for National Statistics (2013) *Disability in England and Wales: 2011 and comparison with 2001*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/disability/articles/disabilityinenglandandwales/2013-01-30 (accessed 01/09/17).

Office for National Statistics (2013) *Inequality in Disability-Free Life Expectancy by Area Deprivation: England: 2003–06 and 2007–10*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/inequalityindisabilityfreelifeexpectancybyareadeprivationengland/2013-07-25 (accessed 24/05/18).

Office for National Statistics (2014) *Overcrowding and Under-occupation in England and Wales*. Available at: http://webarchive.nationalarchives.gov.uk/20160107121343/http://www.ons.gov.uk/ons/dcp171776_360082.pdf (accessed 09/08/17).

Office for National Statistics (2016) *Super Output Area (SOA)*. Available at: http://webarchive.nationalarchives.gov.uk/20160106001702/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html (accessed 01/09/17).

Office for National Statistics (2018) *Health state life expectancies by national deprivation deciles, England and Wales: 2014 to 2016*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthinequalities/bulletins/healthstatelifeexpectanciesbyindexofmultipledeprivationimd/englandandwales2014to2016 (accessed 24/05/18).

Office for National Statistics (no date) *The National Statistics Socio-economic Classification (NS-SEC)*. Available at: https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010 (accessed 29/11/16).

Ogilvie D, Fayter D, Petticrew M, et al. (2008) The harvest plot: A method for synthesising evidence about the differential effects of interventions. *BMC Medical Research Methodology*, 8:8.

Olowokure B, Hawker J, Weinberg J, et al. (1999) Deprivation and hospital admission for infectious intestinal diseases. *Lancet*, 353(9155):807-8.

Omran AR (2005) The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. *The Milbank Quarterly*, 83(4):731–57.

Onozuka D (2014) Effect of non-stationary climate on infectious gastroenteritis transmission in Japan. *Scientific Reports*, 4:5157.

Organisation for Economic Co-operation and Development (no date) *What we do and how*. Available at: http://www.oecd.org/about/whatwedoandhow/ (accessed 26/11/17).

Osborne JW (2013) *Best Practices in Data Cleaning*. Thousand Oaks: Sage.

Özkan S, Tüzün H, Görer N, et al. (2007) Water usage habits and the incidence of diarrhea in rural Ankara, Turkey. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101(11):1131–5.

Özmert EN, Kilic M, Yurdakök K (2008) Environmental tobacco smoke: Is it a risk factor for diarrhea in 6-18 months old infants? *Central European Journal of Public Health*, 16(2):85–6.

Pardo-Crespo MR, Narla NP, Williams AR, et al. (2013) Comparison of individual-level versus area-level socioeconomic measures in assessing health outcomes of children in Olmsted County, Minnesota. *Journal of Epidemiology and Community Health*, 67(4):305–10.

Patterson R, Risby A, Chan M-Y (2012) Consumption of takeaway and fast food in a deprived inner London Borough: are they associated with childhood obesity? *BMJ Open*, 2:e000402.

Pearce N (2000) The ecological fallacy strikes back. *Journal of Epidemiology & Community Health*, 54:326–7.

Pelander T, Leino-Kilpi H (2010) Children's best and worst experiences during hospitalisation. *Scand J Caring Sci*, 24(4):726–33.

PennState (no date) *8.4 - The Proportional-Odds Cumulative Logit Model*. Available at: https://onlinecourses.science.psu.edu/stat504/node/176 (accessed 20/07/17).

Peters JL, Sutton AJ, Jones DR, et al. (2008) Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol*, 61:991–6.

Petersdorf RG (1978) The doctors' dilemma. *N Engl J Med*, 299:628–34.

Phillips G, Tam CC, Rodrigues LC, et al. (2011) Risk factors for symptomatic and asymptomatic norovirus infection in the community. *Epidemiology and Infection*, 139(11):1676–86.

PHLS Advisory Committee on Gastrointestinal Infections (2004) Preventing person-to-person spread following gastrointestinal infections: guidelines for public health physicians and environmental health officers. *Commun Dis Public Health*, 7(4):362–84.

Pickett KE, Pearl M (2001) Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology & Community Health*, 55:111–22.

Pires SM, Fischer-Walker CL, Lanata CF, et al. (2015) Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food. *PLoS ONE*, 10(12):e0142927.

Pockett RD, Adlard N, Carroll S, et al. (2011) Paediatric hospital admissions for rotavirus gastroenteritis and infectious gastroenteritis of all causes in England: an analysis of correlation with deprivation. *Curr Med Res Opin*, 27:777–84.

Polgar S, Thomas SA (2008) *Introduction to Research in the Health Sciences* (5th edition). Edinburgh: Churchill Livingstone Elsevier.

Pollard CM, Meng X, Williamson S, et al. (2014) Eating out is associated with self-reported food poisoning: a Western Australia population perspective, 1998 to 2009. *Public Health Nutrition*, 17(10):2270–7.

Pope D (2015) *Introduction to systematic reviews* [lecture], An Introduction to Systematic Reviews and Meta-analysis. 24 September, University of Liverpool.

Power C, Kuh D (2006) 'Life course development of unequal health', in Siegrist J, Marmot M (eds.) *Social Inequalities in Health: New evidence and policy implications*. New York: Oxford University Press: 27–55.

Public Health England (2013) *PHE Gastrointestinal Infections Data: Summary of eFOSS data, 2013*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/426019/eFOSS_surveillance_tables_for_Web.pdf (accessed 24/04/17).

Public Health England (2014) *Vero cytotoxin-producing Escherichia coli (VTEC): guidance, data and analysis*. Available at: https://www.gov.uk/government/collections/vero-cytotoxin-producing-escherichia-coli-vtec-guidance-data-and-analysis (accessed 24/04/17).

Public Health England (2015) *Rotavirus infant immunisation programme 2014/15: Vaccine uptake report on the temporary sentinel data collection for England*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/440456/RotavirusGatewayFinalVersion.pdf (accessed 09/08/17).

Public Health England (2017a) *Campylobacter: guidance, data and analysis*. Available at: https://www.gov.uk/government/collections/campylobacter-guidance-data-and-analysis (accessed 24/04/17).

Public Health England (2017b) *Salmonella: guidance, data and analysis*. Available at: https://www.gov.uk/government/collections/salmonella-guidance-data-and-analysis (accessed 24/04/17).

Public Health England (2017c) *Shigella: guidance, data and analysis*. Available at: https://www.gov.uk/government/collections/shigella-guidance-data-and-analysis (accessed 24/04/17).

Public Health England (2017d) *Rotavirus data 2006 to 2015: November 2016*. London: Public Health England. Available at: https://www.gov.uk/government/publications/rotavirus-laboratory-confirmed-cases-of-rotavirus-infections-in-england-and-wales (accessed 24/04/17).

Public Health England (2017e) *Norovirus: guidance, data and analysis*. Available at: https://www.gov.uk/government/collections/norovirus-guidance-data-and-analysis (accessed 24/04/17).

Public Health England (2017f) *Cryptosporidiosis: guidance, data and analysis*. Available at: https://www.gov.uk/government/collections/cryptosporidiosis-guidance-data-and-analysis (accessed 24/04/17).

Public Health England (2017g) *Giardia: guidance and data*. Available at: https://www.gov.uk/guidance/giardia (accessed 24/04/17).

Purdy S (2010) *Avoiding hospital admissions: What does the research evidence say?* Available at: https://www.kingsfund.org.uk/sites/default/files/Avoiding-Hospital-Admissions-Sarah-Purdy-December2010.pdf (accessed 14/09/17).

Purdy S, Griffin T, Salisbury C, et al. (2011) Emergency respiratory admissions: influence of practice, population and hospital factors. *Journal of Health Services Research & Policy*, 16:133–140.

Purdy S, Griffin T, Salisbury C, Sharp D (2009) Ambulatory care sensitive conditions: terminology and disease coding need to be more specific to aid policy makers and clinicians. *Public Health*, 123(2):169–73.

Puzzolo E, Stanistreet D, Pope D, et al. (2013) *Factors influencing the large-scale uptake by households of cleaner and more efficient household energy technologies*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Quigley MA, Cumberland P, Cowden JM, et al. (2006) How protective is breast feeding against diarrhoeal disease in infants in 1990s England? A case-control study. *Archives of Disease in Childhood*, 91(3):245–50.

Ramsay SE, Morris RW, Whincup PH, et al. (2014) Time trends in socioeconomic inequalities in cancer mortality: results from a 35 year prospective study in British men. *BMC Cancer*, 14:474.

Rechel B, Mladovsky P, Ingleby D, et al. (2013) Migration and health in an increasingly diverse Europe. *Lancet*, 381(9873):1235-45.

Rehfuess E, Puzzolo E, Stanistreet S, et al. (2014) Enablers and barriers to large-scale uptake of improved solid fuel stoves: a systematic review. *Environ Health Perspect*, 122:120–30.

Reichenheim ME, Coutinho ES (2010) Measures and models for causal inference in cross-sectional studies: arguments for the appropriateness of the prevalence odds ratio and related logistic regression. *BMC Medical Research Methodology*, 10:66.

Richards HM, Reid ME, Watt GCM (2002) Socioeconomic variations in responses to chest pain: qualitative study. *BMJ*, 324:1308.

Richardson WS, Wilson MC, Nishikawa J, et al. (1995) The well-built clinical question: a key to evidence-based decisions [Editorial]. *ACP J Club*, 123(3):A12–3.

Ripley B (2017) *Package 'MASS'*. Available at: https://cran.r-project.org/web/packages/MASS/MASS.pdf (accessed 10/07/17).

Roberts JA, Cumberland P, Sockett PN, et al. (2003) The study of infectious intestinal disease in England: socio-economic impact. *Epidemiol Infect*, 130:1–11.

Rodrigues LC, Cowden JM, Wheeler JG, et al. (2001) The study of infectious intestinal disease in England: Risk factors for cases of infectious intestinal disease with Campylobacter jejuni infection. *Epidemiology and Infection*, 127(2):185–93.

Roland M, Guthrie B (2016) Quality and Outcomes Framework: what have we learnt? *BMJ*, 354:i4060.

Rose D, Pevalin DJ, O'Reilly K (2005) *The National Statistics Socio-economic Classification: Origins, Development and Use*. Basingstoke: Palgrave Macmillan.

Rose TC, Adams NL, Barr B, et al. (2017) Socioeconomic status is associated with symptom severity and sickness absence in people with infectious intestinal disease in the UK. *BMC Infect Dis*, 17:447.

Ross CE, Mirowsky J (2008) Neighborhood socioeconomic status and health: context or composition? *City and Community*, 7(2):163–79.

Ross T (2012) *A survival guide for health research methods*. Maidenhead: Open University Press/McGraw-Hill Education.

Rotheram R (2017) *Conversation with Tanith Rose*, 12 October.

Rougeaux E, Hope S, Law C, et al. (2017) Have health inequalities changed during childhood in the New Labour generation? Findings from the UK Millennium Cohort Study. *BMJ Open*, 7:e012868.

Rubin DB (1976) Inference and missing data. *Biometrika*, 63:581–92.

Rutherford C, Hill J, Sharp H, et al. (2017) Assessing the impact of childhood socio-economic conditions on child mental health: findings from the Wirral Child Health and Development study. *J Epidemiol Community Health*, 71:A11–A12.

Salleh MR (2008) Life Event, Stress and Illness. *The Malaysian Journal of Medical Sciences*, 15(4):9–18.

Scallan E, Fitzgerald M, Collins C, et al. (2004) Acute gastroenteritis in northern Ireland and the Republic of Ireland: a telephone survey. *Commun Dis Public Health*, 7(1):61–7.

Scallan E, Hoekstra RM, Angulo FJ, et al. (2011) Foodborne Illness Acquired in the United States–Major Pathogens. *Emerging Infectious Diseases*, 17(1):7–15.

Scallan E, Jones TF, Cronquist A, et al. (2006) Factors Associated with Seeking Medical Care and Submitting a Stool Sample in Estimating the Burden of Foodborne Illness. *Foodborne Pathogens and Disease*, 3(4):432–8.

Schmid D, Kuo HW, Simons E, et al. (2014) All-cause mortality in hospitalized patients with infectious diarrhea: Clostridium difficile versus other enteric pathogens in Austria from 2008 to 2010. *J Infect Public Health*, 7(2):133–44.

Schmid-Hempel P, Frank SA (2007) Pathogenesis, virulence, and infective dose. *PLoS Pathog*, 3(10):e147.

Seaton AV (1992) Social stratification in tourism choice and experience since the war. *Tourism Management*, 13:106–11.

Segerstrom SC, Miller GE (2004) Psychological Stress and the Human Immune System: A Meta-Analytic Study of 30 Years of Inquiry. *Psychol Bull*, 130(4):601–630.

Sellström E, Bremberg S (2006) The significance of neighbourhood context to child and adolescent health and well-being: a systematic review of multilevel studies. *Scand J Public Health*, 34(5):544–54.

Seo JY, Choi BY, Ki M, et al. (2013) Risk factors for acute hepatitis A infection in Korea in 2007 and 2009: A case-control study. *Journal of Korean Medical Science*, 28:908–14.

Sethi D, Cumberland P, Hudson MJ, et al. (2001) A study of infectious intestinal disease in England: risk factors associated with group A rotavirus in children. *Epidemiology and Infection*, 126(1):63–70.

Shavers VL (2007) Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, 99(9):1013–23.

Sheringham J, Asaria M, Barratt H, et al. (2017) Are some areas more equal than others? Socioeconomic inequality in potentially avoidable emergency hospital admissions within English local authority areas. *J Health Serv Res Policy*, 22(2):83–90.

Siddique J, Harel O, Crespi CM (2012) Addressing Missing Data Mechanism Uncertainty using Multiple-Model Multiple Imputation: Application to a Longitudinal Clinical Trial. *The Annals of Applied Statistics*, 6(4):1814–37.

Siegrist J, Marmot M (2006) 'Introduction', in Siegrist J, Marmot M (eds.) *Social Inequalities in Health: New evidence and policy implications*. New York: Oxford University Press: 1–27.

Simmons K, Gambhir M, Leon J, et al. (2013). Duration of Immunity to Norovirus Gastroenteritis. *Emerging Infectious Diseases*, 19(8):1260–7.

Simon KA, Hollander GA, McMichael A (2015) Evolution of the immune system in humans from infancy to old age. *Proc Biol Sci*, 282(1821):20143085.

Simonsen J, Frisch M, Ethelberg S (2008) Socioeconomic Risk Factors for Bacterial Gastrointestinal Infections. *Epidemiology*, 19:282–290.

Smith KE, Bambra C, Hill S (2016) 'Background and introduction: UK experiences of health inequalities', in Smith KE, Bambra C, Hill S (eds.) *Health Inequalities: Critical Perspectives*. New York: Oxford University Press: 1–22.

Smith S, Elliot AJ, Mallaghan C, et al. (2010) Value of syndromic surveillance in monitoring a focal waterborne outbreak due to an unusual Cryptosporidium genotype in Northamptonshire, United Kingdom, June – July 2008. *Euro Surveill*, 15(33):19643.

Smith T, Noble M, Noble S, et al. (2015) The English Indices of Deprivation 2015: Research report. Department for Communities and Local Government. Available at: https://www.gov.uk/government/publications/english-indices-of-deprivation-2015-research-report (accessed 31/05/18).

Snel SJ, Baker MG, Kamalesh V, et al. (2009) A tale of two parasites: the comparative epidemiology of cryptosporidiosis and giardiasis. *Epidemiol Infect*, 137(11):1641–50.

Solari CD, Mare RD (2012) Housing crowding effects on children's wellbeing. *Social Science Research*, 41(2):464–76.

Sommer I, Griebler U, Mahlknecht P, et al. (2015) Socioeconomic inequalities in non-communicable diseases and their risk factors: an overview of systematic reviews. *BMC Public Health*, 15:914.

Sopori M (2002) Effects of cigarette smoke on the immune system. *Nature Reviews Immunology*, 2(5):372–7.

Spellberg B, Taylor-Blake B (2013) On the exoneration of Dr. William H. Stewart: debunking an urban legend. *Infectious Diseases of Poverty*, 2:3.

Spencer SE, Marshall J, Pirie R, et al. (2012) The spatial and temporal determinants of campylobacteriosis notifications in New Zealand, 2001-2007. *Epidemiology and Infection*, 140(9):1663–77.

Stafford M, Cummins S, Macintyre S, et al. (2005) Gender differences in the associations between health and neighbourhood environment. *Soc Sci Med*, 60(8):1681-92.

Stämpfli MR, Anderson GP (2009) How cigarette smoke skews immune responses to promote infection, lung disease and cancer. *Nat Rev Immunol*, 9(5):377–84.

Stephens RW (2012) 'Medicaid', in Loue S (ed.) *Mental Health Practitioner's Guide to HIV/AIDS*. New York: Springer: 297–300.

Steptoe A (2006) 'Psychobiological processes linking socio-economic position and health', in Siegrist J, Marmot M (eds.) *Social Inequalities in Health: New evidence and policy implications*. New York: Oxford University Press: 101–27.

Steptoe A, Feldman PJ (2001) Neighborhood problems as sources of chronic stress: development of a measure of neighborhood problems, and associations with socioeconomic status and health. *Ann Behav Med*, 23(3):177–85.

Sterne JAC, Egger M, Moher D, et al. (2011b) 'Chapter 10: Addressing reporting biases', in Higgins JPT, Green S (eds.) *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available at: www.handbook.cochrane.org (accessed 26/11/17).

Sterne JAC, Sutton AJ, Ioannidis JPA, et al. (2011a) Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343:d4002.

Sterne JAC, White IR, Carlin JB, et al. (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* , 338:b2393.

Stoltzfus JC (2011) Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, 18:1099–104.

Strand TA, Sharma PR, Gjessing HK, et al. (2012) Risk Factors for Extended Duration of Acute Diarrhea in Young Children. *PLoS ONE*, 7(5):e36436.

Stuebe A (2009) The Risks of Not Breastfeeding for Mothers and Infants. *Rev Obstet Gynecol*, 2(4):222–31.

Subramanian SV, Kawachi I (2006) Whose health is affected by income inequality? A multilevel interaction analysis of contemporaneous and lagged effects of state income inequality on individual self-rated health in the United States. *Health Place*, 12:141–156.

Swanson V, Hannula L, Eriksson L, et al. (2017) 'Both parents should care for babies': A cross-sectional, cross-cultural comparison of adolescents' breastfeeding intentions, and the influence of shared-parenting beliefs. *BMC Pregnancy Childbirth*, 17(1):204.

Tam CC, O'Brien SJ, Tompkins DS, et al. (2012c) Changes in Causes of Acute Gastroenteritis in the United Kingdom Over 15 Years: Microbiologic Findings From 2 Prospective, Population-Based Studies of Infectious Intestinal Disease. *Clin Infect Dis*, 54:1275–86.

Tam CC, Rodrigues LC, O'Brien SJ (2003) The study of infectious intestinal disease in England: what risk factors for presentation to general practice tell us about potential for selection bias in case-control studies of reported cases of diarrhoea. *Int J Epidemiol*, 32:99–105.

Tam CC, Rodrigues LC, Viviani L, et al. (2012a) Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut,* 61:69–77.

Tam CC, Viviani L, Adak B, et al. (2012b) *The second study of infectious intestinal disease in the community (IID2 Study): Final report.* Available at: https://www.food.gov.uk/science/research/foodborneillness/b14programme/b14projlist/b18021 (accessed 28/11/16).

Tam CC, Viviani L, Rodrigues LC, et al. (2013) The second study of infectious intestinal disease (IID2): increased rates of recurrent diarrhoea in individuals aged 65 years and above. *BMC Public Health*, 13:739.

Tanne JH (2007) US needs universal access to health care, American College of Physicians says. *BMJ*, 335(7632):1228.

Tarr PI, Gordon CA, Chandler WL (2005) Shiga-toxin-producing Escherichia coli and haemolytic uraemic syndrome. *Lancet*, 365:1073–86.

Taylor-Robinson DC (2017) *Email communication with Tanith Rose*, 10 November.

Taylor-Robinson DC, Wickham S, Campbell M, et al. (2017) Are social inequalities in early childhood smoking initiation explained by exposure to adult smoking? Findings from the UK Millennium Cohort Study. *PLoS ONE*, 12(6):e0178633.

Teschke K, Bellack N, Shen H, et al. (2010) Water and sewage systems, socio-demographics, and duration of residence associated with endemic intestinal infectious diseases: a cohort study. *BMC Public Health*, 10:767.

Thabane M, Kottachchi DT, Marshall JK (2007) Systematic review and meta-analysis: the incidence and prognosis of post-infectious irritable bowel syndrome. *Alimentary Pharmacology & Therapeutics*, 26:535–44.

Thabane M, Marshall JK (2009) Post-infectious irritable bowel syndrome. World J *Gastroenterol*, 15:3591–6.

The metafor Package: a meta-analysis package for R (2016) *Contour-enhanced funnel plot*. Available at: http://www.metafor-project.org/doku.php/plots:contour_enhanced_funnel_plot (accessed 23/06/17).

Thesaurus.com (no date) *Roget's Thesaurus Alphabetical Index*. Available at: http://www.thesaurus.com/Roget-Alpha-Index.html (accessed 29/11/16).

Thompson SG, Higgins JPT (2002) How should meta-regression analyses be undertaken and interpreted? *Statist Med*, 21:1559–73.

Thompson SG, Sharp SJ (1999) Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, 18:2693–708.

Thrane N, Olesen C, Schønheyder HC, et al. (2003) Socioeconomic factors and prescription of antibiotics in 0- to 2-year-old Danish children. *J Antimicrob Chemother*, 51(3):683-9.

Tian I, Dixon A and Gao H (2012) *Data Briefing – Emergency hospital admissions for ambulatory care-sensitive conditions: identifying the potential for reductions*. London: The King's Fund.

Todkill D, Elliot AJ, Morbey R, et al. (2016) What is the utility of using syndromic surveillance systems during large subnational infectious gastrointestinal disease outbreaks? An observational study using case studies from the past 5 years in England. *Epidemiol Infect*, 144(11):2241–50.

Turin CG, Ochoa T J (2014) The Role of Maternal Breast Milk in Preventing Infantile Diarrhea in the Developing World. *Current Tropical Medicine Reports*, 1(2):97–105.

Turkish Ministry of Health (1995) Turkey 1993: results from the demographic and health survey. *Studies in Family Planning*, 26(3):180–4.

UCLA: Statistical Consulting Group (no date-a) *Ordinal Logistic Regression: R Data Analysis Examples.* Available at: https://stats.idre.ucla.edu/r/dae/ordinal-logistic-regression/ (accessed 07/07/17).

UCLA: Statistical Consulting Group (no date-b) *Negative Binomial Regression: R Data Analysis Examples.* Available at: https://stats.idre.ucla.edu/r/dae/negative-binomial-regression/ (accessed 07/07/17).

UCLA: Statistical Consulting Group (no date-c) *Mixed Effects Logistic Regression: R Data Analysis Examples*. Available at: https://stats.idre.ucla.edu/r/dae/mixed-effects-logistic-regression/ (accessed 10/07/17).

United Nations Children's Fund (2012) *Pneumonia and diarrhoea: Tackling the deadliest diseases for the world's poorest children*. New York: United Nations Children's Fund.

United Nations Children's Fund, World Health Organization (2009) *Diarrhoea: Why children are still dying and what can be done*. New York: United Nations Children's Fund.

United Nations Development Programme (no date) *Human Development Index (HDI)*. Available at: http://hdr.undp.org/en/content/human-development-index-hdi (accessed 29/11/16).

United Nations General Assembly (1948) *Universal declaration of human rights* (217 [III] A). Paris. Available at: http://www.un.org/en/universal-declaration-human-rights/ (accessed 25/04/17).

University of Liverpool (2016) *Library: Health Sciences: Home*. Available at: http://libguides.liverpool.ac.uk/friendly.php?s=health-sciences (accessed 29/11/16).

Van Buuren S, Groothuis-Oudshoorn K (2011) mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3).

Van Cauteren D, De Valk H, Vaux S, et al. (2012) Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiology and Infection*, 140(4):697–705.

Van der Heide I, Wang J, Droomers M, et al. (2013) The Relationship Between Health, Education, and Health Literacy: Results From the Dutch Adult Literacy and Life Skills Survey. *Journal of Health Communication*, 18(Suppl 1):172–84.

Van Lenthe F (2006) 'Aggregate deprivation and effects on health', in Siegrist J, Marmot M (eds.) *Social Inequalities in Health: New evidence and policy implications*. New York: Oxford University Press: 167–93.

van Loenen T, van den Berg MJ, Westert GP, et al. (2014) Organizational aspects of primary care related to avoidable hospitalization: a systematic review. *Fam Pract*, 31(5):502–16.

van Rossum CTM, Shipley MJ, van de Mheen H, et al. (2000) Employment grade differences in cause specific mortality. A 25-year follow-up of civil servants from the first Whitehall study. *Journal of Epidemiology and Community Health*, 54(3):178–84.

Viviani L, van der Es M, Irvine L, et al. (2016) Estimating the Incidence of Acute Infectious Intestinal Disease in the Community in the UK: A Retrospective Telephone Survey. *PLoS ONE*, 11:e0146171.

von Hippel PT (2007) Regression With Missing Y's: An Improved Strategy for Analyzing Multiply Imputed Data. *Sociological Methodology*, 37:83-117.

Voss PH, Rehfuess EA (2013) Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. *J Epidemiol Community Health*, 67:98–104.

Weisberg S (2005) *Applied Linear Regression* (3rd edition). Hoboken: John Wiley & Sons.

Weisent J, Rohrbach B, Dunn JR, et al. (2012) Socioeconomic determinants of geographic disparities in campylobacteriosis risk: a comparison of global and local modeling approaches. *Int J Health Geogr*, 11(1):45.

Wheeler JG, Sethi D, Cowden JM, et al. (1999) Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *BMJ,* 318:1046–1050.

Whitehead M (1991) The concepts and principles of equity and health. *Health Promot Int*, 6(3):217–28.

Whitehead M (1992) 'The health divide', in Townsend P, Whitehead M, Davidson N (eds.) *Inequalities in Health: The Black Report and The Health Divide* (2nd edition). London: Penguin Books: 215–450.

Whitehead M (2007) A typology of actions to tackle social inequalities in health. *Journal of Epidemiology and Community Health*, 61(6):473–8.

Whitehead M, Bambra C, Barr B, et al. (2014) *Due North: Report of the Independent Inquiry on Health Equity for the North of England*. Liverpool: University of Liverpool and Centre for Local Economic Strategies.

Whitehead M, Dahlgren G (2007) *Concepts and principles for tackling social inequities in health: Levelling up Part 1.* Copenhagen: World Health Organisation Regional Office for Europe. Available at: http://www.enothe.eu/cop/docs/concepts_and_principles.pdf (accessed 17/01/17).

Whitehead M, Hanratty B, Burström B (2009) 'Unequal consequences of ill health: researching the role of public policy', in Graham H (ed.) *Understanding Health Inequalities* (2nd edition). Maidenhead: Open University Press/McGraw-Hill Education: 181–99.

Whitehead M, Pennington A, Orton L, et al. (2016) How could differences in 'control over destiny' lead to socio-economic inequalities in health? A synthesis of theories and pathways in the living environment. *Health and Place*, 39:51–61.

Whitney BM, Mainero C, Humes E, et al. (2015) Socioeconomic status and foodborne pathogens in Connecticut, USA, 2000–2011. *Emerging Infectious Diseases*, 21(9):1617–24.

Wickham S, Anwar E, Barr B, et al. (2016) Poverty and child health in the UK: using evidence for action. *Archives of Disease in Childhood*, 101:759–66.

Wikswo ME, Hall AJ (2014) Outbreaks of Acute Gastroenteritis Transmitted by Person-to-Person Contact—United States, 2009–2010. *Am J Public Health*, 104(11):E13–E4.

Wilking H, Höhle M, Velasco E, et al. (2012) Ecological analysis of social risk factors for Rotavirus infections in Berlin, Germany, 2007–2009. *Int J Health Geogr,* 11:1–12.

Wilking H, Spitznagel H, Werber D, et al. (2013) Acute gastrointestinal illness in adults in Germany: a population-based telephone survey. *Epidemiol Infect*, 141(11):2365–75.

Wilkinson RG (1992) Income distribution and life expectancy. *Br Med J*, 304:165–8.

Wilkinson RG (2005) 'Ourselves and others – for better or worse: social vulnerability and inequality', in Marmot M, Wilkinson RG (eds.) *Social determinants of health* (2nd edition). New York: Oxford University Press: 341–59.

Wilkinson RG, Marmot M (2003) *Social Determinants of Health: The Solid Facts* (2nd edition). Copenhagen: World Health Organization. Available at: http://www.euro.who.int/__data/assets/pdf_file/0005/98438/e81384.pdf (accessed 17/01/17).

Wilkinson RG, Pickett KE (2008) Income Inequality and Socioeconomic Gradients in Mortality. *Am J Public Health*, 98:699–704.

Wilson SE, Deeks SL, Rosella LC (2015) Importance of ICD-10 coding directive change for acute gastroenteritis (unspecified) for rotavirus vaccine impact studies: illustration from a population-based cohort study from Ontario, Canada. *BMC Research Notes*, 8:439.

Wingate D, Phillips SF, Lewis SJ, et al. (2001) Guidelines for adults on self-medication for the treatment of acute diarrhoea. *Alimentary Pharmacology & Therapeutics*, 15(6):773–82.

Wintergerst ES, Maggini S, Hornig DH (2007) Contribution of selected vitamins and trace elements to immune function. *Ann Nutr Metab*, 51(4):301–23.

World Health Organization (1948) *Constitution of the World Health Organization*. Geneva: World Health Organization.

World Health Organization (2004) *Climate change and human health: risks and responses*. Geneva: World Health Organization.

World Health Organization (2005) *The treatment of diarrhoea: a manual for physicians and other senior health workers* (4th edition). Geneva: World Health Organization.

World Health Organization (2006) *Five keys to safer food manual*. Geneva: World Health Organization.

World Health Organization (2013) *Diarrhoeal disease: fact sheet*. Available at: http://www.who.int/mediacentre/factsheets/fs330/en/ (accessed 24/04/17).

World Health Organization (2015) *Health in 2015: from MDGs to SDGs*. Geneva: World Health Organization.

World Health Organization (2016a) *Campylobacter: fact sheet*. Available at: http://www.who.int/mediacentre/factsheets/fs255/en/ (accessed 24/04/17).

World Health Organization (2016b) *E. coli: fact sheet*. Available at: http://www.who.int/mediacentre/factsheets/fs125/en/ (accessed 24/04/17).

World Health Organization (2016c) *Salmonella (non-typhoidal): fact sheet*. Available at: http://www.who.int/mediacentre/factsheets/fs139/en/ (accessed 24/04/17).

World Health Organization (2016d) *Guillain–Barré syndrome: fact sheet*. Available at: http://www.who.int/mediacentre/factsheets/guillain-barre-syndrome/en/ (accessed 24/04/17).

Xu Z, Hu W, Tong S (2015) The geographical co-distribution and socio-ecological drivers of childhood pneumonia and diarrhoea in Queensland, Australia. *Epidemiology and Infection*, 143(5):1096–104.

Yang Q, Wen SW, Dubois L, et al. (2004) Determinants of breast-feeding and weaning in Alberta, Canada. *J Obstet Gynaecol Can*, 26(11):975–81.

Young R, Johnson DR (2010) *Imputing the missing Y's: Implications for survey producers and survey users*. 65th Annual Conference of the American Association for Public Opinion Research, 13-16th May, Chicago.

Yuki N, Hartung HP (2012) Guillain-Barré syndrome. *N Engl J Med*, 366:2294–304.

Zhang J, Yu KF (1998) What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*, 280(19):1690–1.

Zwilling ML (2013) Negative Binomial Regression. *The Mathematica Journal*, 15.